Psychovisual Video Coding using Wavelet Transform

by

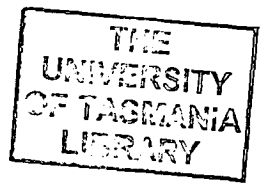
Dadang Gunawan, Ir., M.Eng.

Department of Electrical and Electronic Engineering

Submitted in fulfillment of the requirements
for the degree of
Doctor of Philosophy

University of Tasmania

March 1995



Statement of Originality

This thesis contains no material which has been accepted for the award of any other degree or diploma in any tertiary institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person, except when due reference is made in the text.

Dadang Gunawan

Abstract

Visual communications services are now making a significant impact on modern society. Video conferencing, HDTV and multimedia are just examples where this technology is being used to good effect. Communicating using video signals does, however, require a large volume of data to be transmitted, and even with modern high-bandwidth communication links this can be expensive. This requires the implementation of efficient video coding and compression schemes. This thesis investigates both image and video coding compression schemes and aims to develop a scheme with the highest possible performance.

In image coding there are two main types of compression: statistical and psychovisual. This thesis concentrates on the latter, since it is shown that psychovisual techniques, in general, provide greater levels of compression than statistically based methods. The standardised technique for video coding uses psychovisual compression of the coefficients of the discrete cosine transform (DCT). Despite being an international standard for low bit rate video coding the DCT suffers from a number of drawbacks. Firstly, the psychophysical and psychological models of the human visual system (HVS) are based on a multiresolution approach whereas the basis functions of the DCT are fixed in resolution. Secondly the basis functions of the DCT only possess good localisation properties in the frequency domain and not the spatial domain, a characteristic that blurs edges and discontinuities in an image. By contrast the wavelet transform is a multiresolution approach and its basis functions can possess good localisation properties in both the spatial and frequency domains. Furthermore, due to the excellent localisation properties of the wavelet function most of the transform coefficients are practically zero and the use of wavelet transform can be expected to achieve a higher compression ratio than the DCT. This thesis therefore investigates psychovisual transform coding using the wavelet transform instead of the DCT.

Wavelet basis functions are characterised by a number of parameters, often mutually exclusive. These are spatial compactness, orthonormality, regularity or smoothness and symmetry or anti-symmetry. Since orthonormal and biorthonormal wavelet bases have recently been applied in image coding, a discussion on the design of orthonormal and biorthonormal wavelet bases is presented. By including the properties of the HVS sensitivity function into transform coding using the wavelet transform it is shown that certain design parameters can be relaxed in the construction of an effective image coding scheme [1]. Experimentation using orthonormal wavelets in conjunction with the HVS demonstrated that the length of the filter used to implement the orthonormal wavelet decomposition has only a slight effect on the quality of the reconstructed image. The HVS sensitivity function effectively de-emphasizes the high frequency components thereby relaxing the requirement for smooth wavelets which require longer filters and are more expensive computationally [2].

In image analysis and synthesis the properties of symmetrical or linear phase filters are highly desirable. This can only be achieved by using biorthonormal rather than orthonormal wavelets. Often, the decomposition and reconstruction filters associated with biorthonormal wavelets are of different lengths. Due to the HVS smoothing action, high compression ratios can be achieved at a reasonably low computational cost by using short filters or less regular wavelets on the analysis side [2].

The use of the wavelet transform in image coding can be further divided into two approaches, i.e. a conventional approach and a best basis approach. In the conventional approach, the wavelet basis functions are recursively applied to successively coarser approximation signals, in order to extract the difference in information between consecutive resolutions. The best basis approach involves transforming the image into a set of over-complete basis functions, and permits the choice of a basis function which best suits the image. In terms of computational cost, however, the best basis approach is much more expensive. The comparative computational cost of the conventional and the best basis approaches is $3j + 1$ operations and 4^j operations respectively, where j is the level of decompositions [3].

Finally, in this study, we implement a multiresolution motion compensation technique using a variable block size algorithm for video coding. Due to the large computational cost that renders the best basis algorithm impractical for real-time applications, at least with today's computer processor speeds, the video coding is implemented using the

conventional wavelet approach and not the best basis approach. The most commonly employed motion compensation algorithm is the Block Matching Algorithm (BMA). The performance of this technique is dependent on, among other factors, the size of the fixed blocks employed in this scheme. An alternative to the constant size block matching approach is to use a variable block size. This has the advantage that in areas where the motion is complicated a small block size can be used, and thus the boundaries of moving objects can be more effectively matched [4]. In this study, the proposed variable block matching algorithm as well as the quantization of transform coefficients, is based on the properties of the HVS. In order to do block matching psychovisually, a transform capable of localising an image in both space and frequency is required. This enables the block matching to be performed in the transform domain, after psychovisual thresholding [5]. The wavelet transform is capable of meeting this requirement. By contrast the DCT is not suited to this approach, because any localised motion within the image is lost in the transform. The simulation results presented in this study show that the bit-rates of the proposed scheme are superior compared to the existing schemes.

Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisor Professor D. Thong Nguyen for his valuable guidance, encouragement, helpful discussions and availability for consultation during the years it took to complete my research. I wish also to express my sincere thanks to Dr. Richard G. Lane, for his helpful and expert advice. I would also like to thank all the staff of the Electrical & Electronics Engineering Department at the University of Tasmania, and all my fellow Ph.D students, especially Marc A. Stoksik for his encouragement and proof-reading of the manuscript. Finally, I would like to thank the EMSS - AIDAB and Yayasan Pendidikan Teknik Indonesia for financing this research, especially Mr. Christ Street and Mrs. Robin Bowden for their help throughout the course of my research. Needless to say, very special thanks must be given to my beloved family, my wife and my children.

Glossary

This thesis contains a certain amount of mathematics. The following notations and symbols are employed.

$f(x)$, $f(x,y)$	One and two dimensional signals.
$I(x,y)$	Two-dimensional image.
$\mathbf{Z}$ and $\mathbf{R}$	Set of integers and real numbers.
$\mathbf{L}^2(\mathbf{R})$	Vector space of measurable, square-integrable one-dimensional functions.
$\mathbf{L}^2(\mathbf{R}^2)$	Vector space of measurable, square-integrable two-dimensional functions.
$l^2(\mathbf{Z})$	Vector space of square-summable sequences.
V_j	One-dimensional ladder spaces of j.
$\mathbf{V}_j$	Two-dimensional ladder spaces of j.
W_j	Orthogonal complement of space V_j in V_{j-1}.
$\mathbf{W}_j$	Orthogonal complement of space $\mathbf{V}_j$ in $\mathbf{V}_{j-1}$.
$\phi(x), \Phi(x,y)$	One and two dimensional scaling functions.
$\psi(x), \Psi(x,y)$	One and two dimensional wavelet functions.
g	Bandpass filter in time domain.
G	Bandpass filter in frequency domain.
h	Lowpass filter in time domain.
H	Lowpass filter in frequency domain.
T_b	Number of bits.
C_t	Channel capacity.
F_r	Frame rates.
I_n	n^{th} sample of the sequence input.

$E[x]$ or $E(x)$	Expected value of x.
σ^2	variance.
W_L	Code word length.
$P()$	Probability.
$H_e(x)$	Entropy of random variable x.
$\|x\|$	Absolute value of x.
$\langle x, y \rangle$	Inner product of x and y.
$\hat{x}$	Fourier transform of x.
$\bar{x}$, x^*	Complex conjugate of x.
$\delta_{i,j}$	The Kronecker delta, equal 1 if $i = j$ and 0 otherwise.
$\propto$	Proportional to.
∞	Infinity.
$\in$	A member of.
$\subset$	A subset of.
$\cup$	Union.
$\cap$	Intersection.
$\leftrightarrow$	If and only if.
Σ	Sum.
Π	Product.
$\mathbf{A}$	Matrix realization of operator A.
$\mathbf{I}$	An identity matrix.
$\mathbf{A}^{-1}$	Inverse of the matrix A.
$\mathbf{A}^T$	Matrix transpose of A.
$\otimes$	Tensor product.
$\odot$	Convolution.
$\oplus$	A sum operation.
$\equiv$	Identically equal to.
π	Pi (3.1415...).
i	$\sqrt{-1}$.

The following abbreviations have also been used:

ADPCM	Adaptive Predictive Pulse Code Modulation.
ATM	Asynchronous Transfer Mode.

BISDN	Broadband Integrated Services Digital Network.
BMA	Block Matching Algorithm.
BMA-MC	Block Matching Algorithm-Motion Compensation.
BOS	Beginning Of Subblock.
bpp	Bits per pixel.
bps	Bits per second.
CCF	Cross Correlation Function.
CCIR	Consultative Committee International on Radio
CCITT	Consultative Committee International on Telegraph and Telephone.
CIF	Common Intermediate Format.
CWT	Continuous Wavelet Transform.
cpd	Cycles per degree.
dB	Decibels.
DCT	Discrete Cosine Transform.
DFT	Discrete Fourier Transform.
DM	Delta Modulation.
DPCM	Differential Pulse Code Modulation.
DWT	Discrete Wavelet Transform.
FFT	Fast Fourier Transform.
FIR	Finite Impulse Response.
HDTV	High Definition Television.
HVS	Human Visual System.
Hz	Hertz.
IWT	Inverse Wavelet Transform.
KLT	Karhunen-Loeve Transform.
Kbits	Kilo (10^3) bits.
MAD	Mean Absolute Difference.
Mbps	Mega (10^6) bits per second.
MC	Motion Compensation.
MSE	Mean Square Error.
MSPE	Mean Square Prediction Error.
NTSC	National Television System Committee.
PAL	Phase Alternation Line.
PCM	Pulse Code Modulation.

PSNR	Peak Signal-to-Noise Ratio.
QCIF	Quadrature Common Intermediate Format.
QMF	Quadrature Mirror Filter.
RLP	Run-Length Prefix.
SECAM	Sequentiel Couleur Avec Memoire.
SNR	Signal-to-Noise Ratio.
STFT	Short-Time Fourier Transforms.
WHT	Walsh-Hadamard Transform.
WPSNR	Weighted Peak Signal-to-Noise Ratio.
WT	Wavelet Transform.

Preface

The standardized techniques for image and video coding use a psychovisual compression of the coefficients of the discrete cosine transform (DCT). Despite being an international standard for low bit rate video coding the DCT suffers from a number of drawbacks. Firstly, the psychophysical and psychological models of the human visual system (HVS) are based on a multiresolution approach whereas the DCT basis functions are fixed in resolution. Secondly the basis functions of the DCT only possess good localisation properties in the frequency domain and not the spatial domain, a characteristic that blurs edges and discontinuities in an image. By contrast the wavelet transform is a multiresolution approach and its basis functions can possess good localisation properties in both the spatial and frequency domains. Furthermore, due to the excellent localisation properties of the wavelet function most of the transform coefficients are practically zero and the use of the wavelet transform can be expected to achieve a higher compression ratio than the DCT. The original purpose of this research was to design and develop a high performance image and video scheme using a technique that involves the psychovisual coding of wavelet transform coefficients.

Thesis Organization

This thesis is organized into six chapters. Chapter 1 gives a brief introduction of visual communications and their source format. Chapter 2 provides a detailed summary of statistical and psychovisual image coding as well as describing a compression technique for single frame images and sequence of images. Chapter 3 outlines the design of wavelet bases and their properties. Chapter 4 outlines and examines the proposed image compression scheme for single frame images. A comparison of the performances for all

the coding schemes using various wavelet bases are also investigated in this chapter. Chapter 5 outlines the extension of the proposed scheme for a sequence of images. A comparison of the performances of the proposed scheme and existing schemes is then presented. Finally, Chapter 6 contains a summary of the major results and suggestions for future research.

Supporting Publications

This research resulted in a number of journal and conference publications during the course of this study. There are listed below :

1. D. Gunawan and D. T. Nguyen, "Psychovisual Image Coding Using Wavelet Transform", In *Australian Journal of Intelligent Information Processing Systems, Autumn Issue Vol. 2, No. 1, pp. 45 - 52, March 1995.*

2. D. Gunawan, R. G. Lane and D. T. Nguyen, "Adaptive Motion-Compensated Interframe Prediction Coding Using Subjective Thresholding of Wavelet Transform Coefficients", In *1994 IEEE Singapore International Conference on Communication Systems, ICCS'94*, pp. 1135 - 1137, Singapore, November 1994,

3. D. T. Nguyen and D. Gunawan, "Wavelets and Wavelets - Design Issues", In *1994 IEEE Singapore International Conference on Communication Systems, ICCS'94*, pp. 188-194, Singapore, November 1994.

4. D. Gunawan and D. T. Nguyen, "Subjective Coding of Wavelet Coefficients for Multiresolution Analysis", In *Proceedings of Image & Vision Computing Conference*, pp. 163-168, Auckland, New Zealand, August 1993.

5. D. L. McLaren, D. Gunawan and R. G. Lane, "Motion Detection and Adaptive Compensation for Efficient Interframe Video Coding", In *The Third International Symposium on Signal Processing and Its Applications (ISSPA'92) Proceedings*, pp. 634-637, Gold Coast, Australia, August 1992.

6. D. Gunawan, D. T. Nguyen and R. G. Lane, "Subjective Coding of Best Basis Wavelet Coefficients for Multiresolution Analysis", Submitted to *Signal Processing : Image Communication*, December 1994.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Nowadays, video applications such as digital laser disc, electronic camera, videophone and video conferencing systems, image and interactive video tools on personal computers and workstations, program delivery using cable and satellite, and high-definition television (HDTV) are available for visual communications. Many of these applications, however, require the use of data compression because visual signals require a large communication bandwidth for transmission and large amounts of computer memory for storage. In order to make the handling of visual signals cost effective it is important that their data be compressed as much as possible. Fortunately, visual signals contain a large amount of statistically and psychovisually redundant information. By removing this unnecessary information, the amount of data necessary to adequately represent an image can be reduced.

The removal of unnecessary information generally can be achieved by using either statistical compression techniques or psychovisual compression techniques. Both techniques result in a loss of information, but in the former the loss may be recovered by signal processing such as filtering and inter or intra-polation. In the latter, information is in fact discarded, but in way that is not perceptible to a human observer. The latter technique offers much greater levels of compression but it is no longer possible to perfectly reconstruct the original image. While the aim in psychovisual coding is to keep these differences at an imperceptible level, psychovisual compression inevitably involves a tradeoff between the quality of the reconstructed image and the compression

rate achieved. This tradeoff can often be assessed using mathematical criteria, although a better assessment is in general provided by a human observer.

The compression techniques can be applied to single or still images and to sequences of video images. There are two main techniques of image compression. The first technique, known as intra-frame coding, relies on removing the spatial redundancies in the images a single frame at a time. Very often these techniques are simply called image compression techniques. The second technique, inter-frame coding, exploits the redundancies caused by temporal correlations as well as spatial correlations in successive video images. This technique holds the promise of a significantly large reduction in the data required to transmit the image sequence.

There are numerous ways to achieve compression in both image compression and video compression techniques. Predictive coding and transform coding are widely used for image compression techniques. Predictive schemes compress each of the picture elements (pixels) by quantizing the difference between a predicted value whose value is based on the previous pixels. Transform coding especially using the Discrete Cosine Transform (DCT) is an international standard and one of the most powerful compression techniques [6]. In successive video images, each picture frame is typically very close to those temporally adjacent to it and each picture usually contains a background and a number objects, which are essentially unchanging from frame to frame. When some of the objects are moving, and assuming that movement in the picture is only a shift or displacement of object position, then the location of a pixel or a group of pixels on the same part of the moving object can be predicted from the previous frame. This technique is called motion compensation.

This thesis is largely concerned with psychovisual compression techniques. The applications of intra-frame and inter-frame coding in conjunction with transform coding and psychovisual compression are presented in this study.

1.1 Visual Communication

Good visual communication requires an integration of the telecommunication networks. Fortunately, the advances in recent network technology such as the powerful ATM-based Broadband ISDN and extensive optical networks have been conveniently developed for

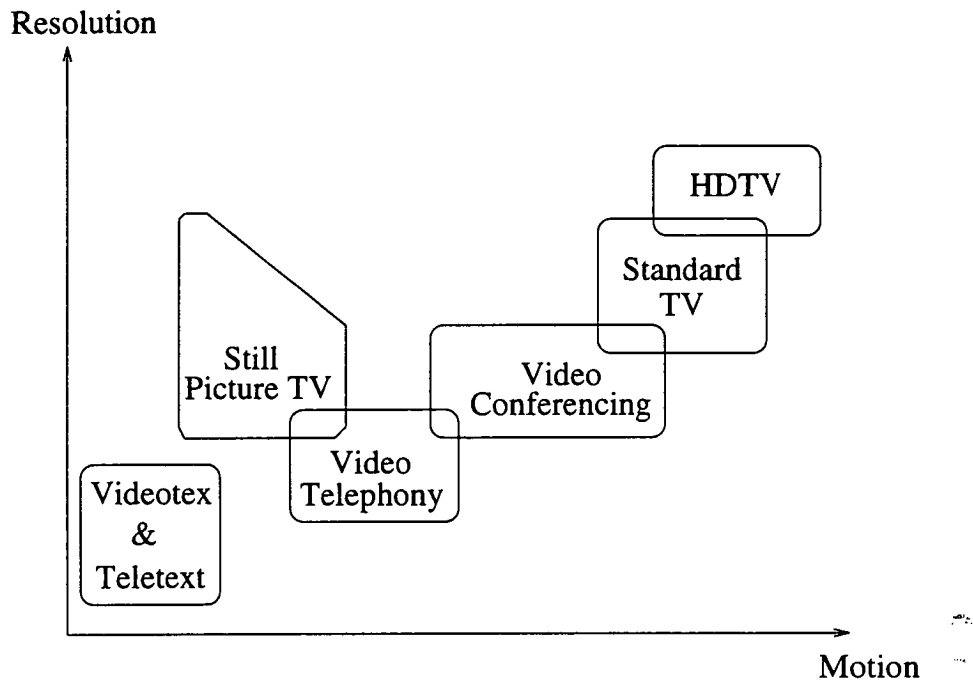


Figure 1.1: Visual services according to resolution and movement.

visual communication services. These visual communication services can be distinguished as "interactive services" and "distribution services" [7]. Interactive services contain conversation services, messaging services and retrieval services. Conversation services provide dialogue communication with real time end-to-end information transfer such as data processing. Messaging services offer user-to-user communication between individual users via storage units with store-and-forward functions. Retrieval services can retrieve information stored in information centres and, in general, be provided for public use. The distribution services, also known as broadcast and narrowcast services, provide the customer with an information flow from a central source. Figure 1.1 illustrates a collection of some of the common video and image-based services in accordance to resolution and movement.

Videotex, Teletext and Still Picture TV

Videotex [8] [9] is based on a combination of television, communications and computer technologies. Videotex is a two way system in which a user can access a remotely located computer from their home, business, or present location and display information

retrieved from the computer on a modified television set or specially designed visual display unit. Videotex services enable information in the forms of words, pictures and graphics to be transmitted and displayed electronically on demand, and with which the user can interact by sending electronic messages or commands in response.

Teletext [9] refers to any broadcasting system which displays selected frames of information as they are being continuously recycled by the originator of the signal. The broadcast signal is generated with the aid of a computer. The information is prepared and stored digitally and is usually broadcast as a portion of the regular television signal. Still picture TV [7, pp. 8-11], often referred to as slow-scan television or freeze frame, was developed for applications where the update of movement is not critical.

Video Conferencing

Video conferencing systems are used for two or more groups of people at widely separated locations who wish to communicate in real time, both visually and orally [10] [11]. The objective of video conferencing is to save time and costs that would be incurred by physically bringing all of the conference participants to the same location and to be a useful alternative form of communication between co-workers, business associates and managerial staff. During the video conference several types of information can be transmitted such as sound, still and moving video, messages, handwriting and facsimile.

Since the service of video conferencing is much more task oriented compared to broadcast television which is much more entertainment oriented, the picture quality constraints are less stringent than for broadcast television [11, pp. 115-120]. The standard transmission format for these interactive services uses Common Intermediate Format (CIF) which is tabulated in Table 1.2. Standardized bit-rates are not sufficient for internationally compatible video conferencing services between countries with different local TV systems. The most useful Integrated Services Digital Networks (ISDN) bit-rates for video conferencing services use 384 Kbps and multiples thereof up to and including the primary rates [11, pp. 115-120] [12].

Video Telephony

Video telephony is a bidirectional service in which speech and moving images are communicated as in normal face-to-face conversation, including facial expressions, posture and gestures [11, pp. 115-120] [13]. The picture information transmitted is sufficient for the adequate representation of smooth motions of persons displayed in head and shoulder view. The video telephone has a mode for normal telephone, and narrow band video telephony can basically be seen as scaled-down video conferencing with limited functionality. The standard transmission format of Quadrature CIF (QCIF) and temporal frequencies below 15 frames-per-second with grey-scale images are commonly used for video telephony.

Multimedia Communication

Recent advances in networking, storage, personal computers, workstations and integrated circuit technology as well as demands for better communication accommodations have fostered tremendous interest in the development of multimedia communication systems [14]. Multimedia communication [15] is the field referring to the representation, storage, retrieval and dissemination of machine-processable information expressed in multimedia, such as text, audio, voice, graphics, image and video in a range of configurations to suit the end user.

Applications in education and training, office and business systems, information and point of sales are the main multimedia services [16]. In all of these different fields, the visual aspects of the multimedia applications are still image, motion video and graphics based.

1.2 Source Format of Video Signals

All video communication services use television signals with different formats. This section describes the video signal source format of several communication systems. Television signals comprise a sequence of still pictures or frames, and each frame consists of two interlaced fields. Two major standards for regional television are the PAL and

NTSC systems. The PAL system is used in Europe (with the exception of France which uses SECAM) and Australia with the main scanning standard of 50 fields-per-second and 625 lines-per-frame. The NTSC system is used in North America and Japan with 60 fields-per-second and 525 lines-per-frame. These standards have been adopted by the CCIR recommendation 601 [17]and are depicted in Table 1.1.

	PAL	NTSC
Parameters	625/50	525/60
sampling frequency	13.5 MHz (luminance Y)	13.5 MHz
	6.75 MHz (chrominance U,V)	6.75 MHz
sampling structure	orthogonal	orthogonal
number of active lines	576	480
number of pixels per active line		
luminance (Y)	720 pixels/line	720 pixels/line
chrominance (U)	360 pixels/line	360 pixels/line
chrominance (V)	360 pixels/line	360 pixels/line
Field frequency	50 Hz	59.94 Hz
quantization	8 bit PCM	8 bit PCM

Table 1.1: Parameters of CCIR 601 Source Format Standard.

The luminance component Y is sampled at 13.5 MHz and the chrominance components U and V at 6.75 MHz. Therefore the number of chrominance pixels is half the number of luminance pixels. The number of bits in one frame for either standard can be computed by

$$T_b = Hor(m) \times Ver(n) \times \text{bpp}$$

where T_b denotes the number of bits in one frame, bpp the number of bits per pixel, $Hor(m)$ the number of horizontal samples and $Ver(n)$ the number of active lines. When dealing with PAL and NTSC systems, the total number of bits becomes

$$\begin{aligned} T_{b\,PAL} &= (720 \times 576 \times 8) + 2(360 \times 576 \times 8) \\ &= 6,635,520\text{bits/frame}, \end{aligned}$$

and

$$\begin{aligned} T_{b\,NTSC} &= (720 \times 480 \times 8) + 2(360 \times 480 \times 8) \\ &= 5,529,600\text{bits/frame}, \end{aligned}$$

respectively.

The overall bit rate is defined by

$$C_t = T_b \times F_r$$

where F_r denotes the frame rates. The frame rate, F_r, for PAL is 25 Hz and for NTSC is approximately 30 Hz. Therefore the channel capacity to transport CCIR rec. 601 signals is $C_{t\ PAL} = 166$ Mbps and $C_{t\ NTSC} = 166$ Mbps respectively.

Common Intermediate Format

Many regions have different national television standards and it is therefore necessary to convert standards for use in inter-regional television broadcasting. The Common Intermediate Format (CIF) was created when an agreement was reached to overcome the differences between inter-regional television standards [7, pp. 31-33]. The NTSC system actually has 480 active lines, whereas the PAL system has 576 active lines. It was agreed internationally that the countries using NTSC would convert the number of lines towards 288 non-interlaced lines and the countries using PAL would convert the number of fields. The basic parameters for CIF are depicted in Table 1.2. A fallback mode is based on one quarter of the CIF, and is denoted by QCIF.

	CIF	QCIF
Number of active lines		
Luminance (Y)	288	144
Chrominance (U, V)	144	72
Number of active pixels per line		
Luminance (Y)	360	180
Chrominance (U, V)	180	90

Table 1.2: Parameters of the Common Intermediate Format

1.3 Structure of the thesis

The main focus of this thesis is the design of psychovisual image and video compression schemes using the wavelet transform. The next two chapters describe the available techniques for image and video coding and the chapter after that introduces the wavelet transform. The last two chapters focus on the design of a psychovisual image compression and video compression technique using the wavelet transform respectively.

Chapter II describes an overview of the techniques available for image and video compression of visual information. The statistical image compression techniques such as PCM, Predictive coding, Transform coding, Pyramid coding and Sub-band coding, as well as psychovisual image coding techniques are presented. Combinations of these techniques are also outlined.

Chapter III describes the wavelet transform which is a useful sub-band coding scheme based on a multiresolution decomposition of signals and images. The wavelet basis functions can be generated using many different functions. We construct, in particular, wavelet bases by using the methodology for generating compactly supported orthonormal and biorthonormal wavelet bases. The properties of such wavelets are also presented. These wavelet bases are then employed for statistical as well as psychovisual image compression.

Chapter IV explores psychovisual image compression techniques using the wavelet transform coefficients. The application of the wavelet transform in signal and image processing can be divided into two schemes: the conventional approach based directly on Mallat's multiresolution wavelet decomposition [18] and the best basis approach [19]. This chapter firstly examines the conventional scheme using both psychovisual compression as well as statistical compression with various wavelet bases. The simulation results of both the statistical and psychovisual schemes are compared for the various wavelet bases. Only the wavelet bases which produced the best result are applied to the best basis approach and used for a comparison with the conventional scheme.

Chapter V extends the psychovisual compression techniques by exploring the video coding of the wavelet transform coefficients. The basic video coding technique of using a motion compensation scheme both with fixed block size and variable block size matching algorithms is presented. Following this a multiresolution motion compensation scheme

is examined, in particular for the conventional wavelet transform scheme. The best basis scheme is not carried out using the multiresolution motion compensation scheme, due to a large computational cost that renders the algorithm impractical for real-time applications.

Chapter VI provides an overall summary of the results presented, as well as suggestions for future research.

Chapter 2

From Image to Video Compression

2.1 Introduction

An image is usually represented by a two-dimensional array of numbers over a rectangular or square lattice. The gray level variations are used to represent the information in the image. In many practical cases, an image is often defined over a 256×256 or 512×512 lattice. Typically each pixel is represented by 8 bits, corresponding to a gray level variation between 0 and 255. Storing a 512×512 image requires approximately 2096 Kbits memory which constitutes a significant quantity of memory. It is therefore desirable to compress the information in the image into a considerably fewer number of bits whilst maintaining the ability to reconstruct the image such that it is close to the original image. Thus an image needs to be compressed for efficient data storage applications or to reduce the bandwidth capacity required to transmit the image.

The applications of image data compression, in general, are primarily in the transmission and storage of information. In transmission, applications such as broadcast television, teleconferencing, videophone, computer-communication, remote sensing via satellite or aircraft, etc., require the compression techniques to be constrained by the need for the real time compression and on-line consideration which tends to severely

limit the size and hardware complexity. In storage applications such as medical images, educational and business documents, etc., the requirements are less stringent because much of the compression processing can be done off-line. However, the decompression or retrieval should still be quick and efficient to minimise the response time [20].

When dealing with images the data compression techniques may be divided into two schemes. There are image compression techniques and video compression techniques. The former techniques exploit the redundancy in the images a single frame at a time, whilst the latter ones exploit the temporal correlations between successive images in the sequence. If the sequence is highly correlated, this technique holds the promise of significantly greater compression in the data when transmitting an image sequence. If there exists no correlation within the various images making up the sequence, the problem of video compression reduces to one of image compression, with each individual image within the sequence treated separately [21].

All images of interest usually contain a considerable amount of statistically and subjectively superfluous information [11, pp. 148-150]. A statistical image compression technique exploits statistical redundancies in the information in the image. This technique reduces the amount of data to be transmitted or to be stored in an image without any information being lost. The alternative is to discard the subjective redundancies in an image, which leads to psychovisual image compression. These psychovisual techniques rely on properties of the Human Visual characteristic System (HVS) to determine which features will not be noticed by a human observer.

In the following section a brief overview of image coding and video coding in terms of compression techniques is provided. Both statistical and psychovisual compression schemes are presented. More general reviews on the subjects can be found in [7] [11] [20] [22] [23] [24] [25].

2.2 Statistical Image Compression Techniques

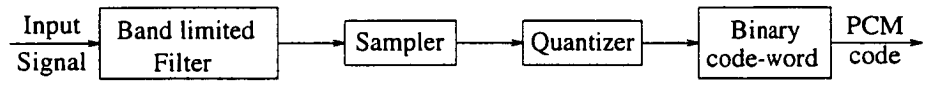
As noted above visual images generally contain large amounts of statistically redundant information. There is normally considerable correlation between adjacent pixels both vertically and horizontally in an image, as well as between pixels in successive frames. In statistical image compression techniques little or no information in the image is

discarded and therefore a perfect or near perfect representation of the image can be reconstructed. In this section we discuss some compression methods which belong to the class of statistical compression techniques.

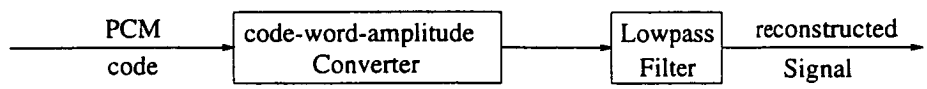
2.2.1 Pulse Code Modulation (PCM)

PCM, also known as analog to digital conversion, is a simple method to represent the discrete amplitudes of signal information into binary code words. It has not been used for television up until 1951, although it was developed in the late 1930's. Since then it has been used as a video digitizing scheme for the purposes of storage and transmission.

The block diagrams of a PCM encoder and decoder are shown in Figure 2.1. The input signal, a one-dimensional raster scanned waveform of the image is firstly band-limited by an anti-aliasing filter and then sampled at the Nyquist sampling rate, where the Nyquist sampling rate is twice the highest frequency in the input signal. The resulting sampled signal is then quantized to 2^N discrete amplitude levels. Each level is represented by a binary code word containing N bits. This scheme is also known as fixed word-length coding. At the decoder, these binary code words are converted to a sequence of discrete amplitude levels which are then lowpass filtered to obtain a reconstruction of the original signal. The number of bits needed to code a pixel using PCM depends on the type of image. In general, 128 or 256 levels (7 or 8 bits) are sufficient for a monochrome image [11, pp. 304-307].



(a) Encoder PCM



(b) Decoder PCM

Figure 2.1: Block diagrams of a PCM encoder and decoder.

In order to obtain a perfectly reconstructed image from the coded data, the number of quantizing levels, 2^N, must be sufficient to ensure that the decoded image is not significantly different from the original image. Thus, in this simple scheme each pixel is represented by N bits.

2.2.2 Predictive Coding

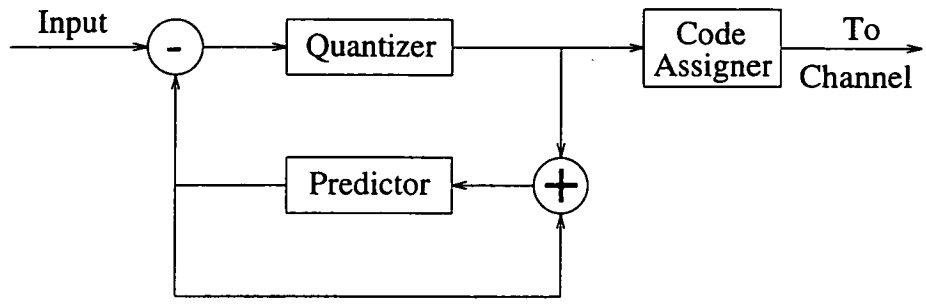
In PCM coding, the successive inputs to the quantizer are treated independently, and thus there is no exploitation of any redundancy present in images. Visual images contain statistical redundancies such as the correlation between adjacent pixels that are spatially close to each other [11, pp. 313-321]. Predictive coding techniques exploit this correlation. The principle of this technique is to predict or to approximate the next sample in order to remove the mutual redundancy between successive samples and quantize only the difference or new information.

The block diagram of a predictive coder and decoder is shown in Figure 2.2. The predictive coder has three basic components; namely, a Predictor, Quantizer and Code Assigner. The basic principle of this technique is to predict the value of the current pixel based on the previously coded pixel that has been transmitted [26]. The prediction error or differential signal is quantized into a set of N discrete amplitude levels, coded using either a fixed or variable word-length code and then transmitted. The quantizer depends on the number of levels of quantization. If the quantizer has only 2 levels, the predictive coding is known as Delta Modulation (DM) [27]. When the quantizer has more than 2 levels the predictive coding is known as Differential Pulse Code Modulation (DPCM). Since, DM uses only 2 quantization levels, the sampling rate has to be several times the Nyquist rate in order to avoid slope overload and get an adequate picture quality [11, pp. 313-321].

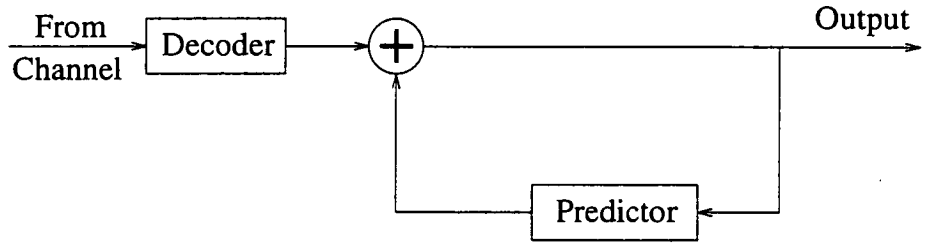
The predictor for DPCM can be either linear or non-linear, depending on whether the prediction is a linear or non-linear combination of previously transmitted values. Linear prediction has been extensively studied [28] [29] and has the general form

$$\check{I}_n = \sum_{p=1}^{P} \alpha_p I_{n-p} \tag{2.1}$$

where I_n is the n^{th} sample of the input sequence, α_p is the p^{th} prediction coefficient



(a) Coder



(b) Decoder

Figure 2.2: Block diagram of an encoder and decoder for Predictive Coding.

and P is the order of prediction. The prediction coefficients $\{\alpha_p\}$ can be obtained by minimising the mean square prediction error (MSPE), $E(I_n - \check{I}_n)^2$. By using the optimum coefficients, the mean square prediction error is given by [11, pp. 313-321]

$$\text{optimum(MSPE)} = \sigma^2 - \sum_{p=1}^{P} \alpha_p dp \tag{2.2}$$

where $d_p = E\{I_n . I_{n-p}\}$ and the pixels are assumed to be identically distributed with zero mean and variance σ^2. The quantizer output is then encoded using either a fixed or variable word-length code. If either the quantizer step size Δ is adapted to the rate of change of the signal or the prediction coefficients $\{\alpha_p\}$ are adapted to the temporal change of the signal, we have an Adaptive Predictive PCM (ADPCM).

2.2.3 Code word assignment and Huffman Coding

In both PCM and DPCM schemes, the output quantized levels are coded by the assignment of code words before being stored or transmitted. In general, the PCM scheme uses constant or fixed length N bit binary words to represent the various signal levels. In practice, however, the probability distribution of the output quantizer levels, especially for DPCM, is highly nonuniform. This naturally lends itself to a representation using code words of variable length [11, pp. 379-380] [30]. The average number of bits per word can be reduced if the output quantized levels having a high probability are assigned short code-words, while the output quantized levels having a lower probability are assigned longer code words. This method is called variable word-length coding or entropy coding [11, pp. 148-150]. Let the output quantized level be denoted by b with probability of occurence $P(b)$. It is assigned a code word of length $W_L(b)$ bits. Then the average code word length $\overline{W}_L$ can be calculated by

$$\overline{W}_L = \sum_b W_L(b)P(b) \text{ [bits/pixel]}. \tag{2.3}$$

The average code word length $\overline{W}_L$ cannot be made arbitrarily small, however, and still must be correctly decoded by a receiver [11, pp. 148-150]. The average code word length $\overline{W}_L$ should satisfy the lower bound which is derived from information theory [31], and is given by

$$H_e(B) \leq \overline{W}_L \tag{2.4}$$

where $H_e(B)$ is the entropy,

$$H_e(B) = -\sum_b P(b)\log_2 P(b) \text{ [bits/pixel]}. \tag{2.5}$$

The most popular entropy coding scheme employs Huffman coding [32], in which the statistical distribution of the output quantizer levels is used to construct a look-up table of optimum variable length code words. In practice, it is found that the performance of the Huffman code is not too sensitive to small changes in the probablity distribution from picture to picture [30], and therefore efficient code word assignment can be provided if codes based on the average distribution of levels over many images are used [11, pp. 379-380].

2.2.4 Transform Coding

The previous coding techniques attempt to reduce the correlation that exists among adjacent image pixel intensities. Transform coding [11] [39] [40] attempts to represent an image by uncorrelated data, using a reversible linear transformation. In the most general form, an image is sampled and subdivided into small square blocks of $N \times N$ pixels. The transformation of one of these subblocks into another domain, usually frequency-related, is denoted by

$$\hat{I}(u,v) = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} I(m,n) f_{ker}(u,v;m,n), \text{ for } u,v;m,n = 0,1,\cdots,N-1 \qquad (2.6)$$

where $f_{ker}(u,v;m,n)$ is the forward transform kernel. The original image subblocks can then be recovered by

$$I(m,n) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \hat{I}(u,v) i_{ker}(u,v;m,n), \text{ for } u,v;m,n = 0,1,\cdots,N-1 \qquad (2.7)$$

where $i_{ker}(u,v;m,n)$ is an inverse transform kernel.

The objective of transform coding is to represent the energy of the original subblock by a small fraction of the transform coefficients, a process called energy compaction. This means that only the transform coefficients that adequately represent the image need to be transmitted.

The block diagram of the transform coding technique is shown in Figure 2.3. The original image is subdivided into small square subblocks and transformed. The non-zero transformed subblock coefficients in which all the energy is contained are quantized and encoded before transmission. DPCM is, in general, used for the coding scheme. At the receiver, the code words are decoded and then dequantized. These received subblocks are then inverse transformed to obtain the reconstructed image.

Once the non-zero transform coefficients are quantized, they then need to be coded into a binary stream before being stored or transmitted. The non-zero coefficients are, in general, coded by a magnitude look-up table and the addresses of the coefficients are coded using a runlength look-up table [36]. The amplitude and run-length look-up

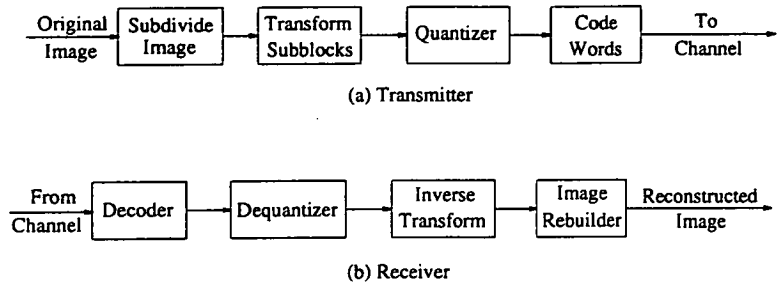


Figure 2.3: Block diagram of Transform Coding technique.

tables are simply Huffman coded. In order to minimise the amount of information required to represent a subblock, the quantized coefficients need to be scanned before coding. There are many ways in which an image can be scanned but the zig-zag scanning method illustrated by Figure 2.4 is the most commonly used and has been accepted as part of an international coding scheme standard [35] [36] [37].

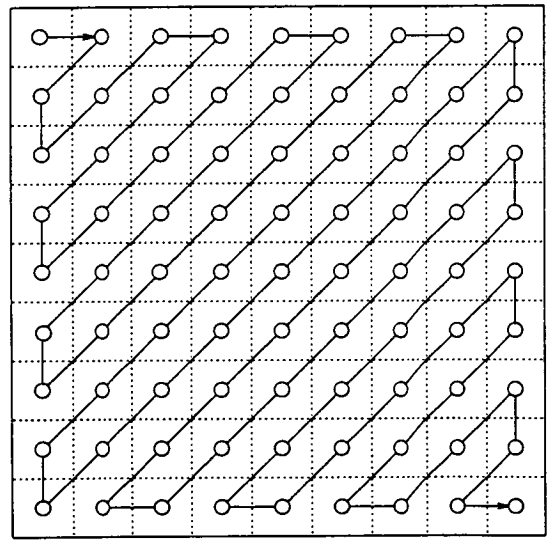


Figure 2.4: Zig-zag scanning of transform coefficients.

The choice of basis kernel in equation (2.6) results in different forms of transform compression. The optimal choice of basis kernel which results in the most uncorrelated transform and greatest image compaction is known as the Karhunen-Loève transform (KLT). However, the transform matrix or the kernel must be derived from the corre-

lation statistics of the original subblocks [11, pp. 392-400], which makes it difficult to implement. Consequently, the KLT is not often used in practice, but rather provides an optimal linear transformation for bench-mark comparison of other techniques. There are many different less optimal linear transforms which produce compactly efficient algorithms and which are independent of the statistics of the input images [11]. Some of the more important transforms are the discrete Fourier transform (DFT) [34], Walsh-Hadamard transform (WLH) [38] and discrete Cosine transform (DCT) [39]. Recently, the wavelet transform (WT) has also be used in transform image coding.

Discrete Fourier Transform

The DFT is a very common separable unitary and symmetric linear transform, which transforms the spatial domain into the frequency domain. The basis kernels of the DFT are the complex exponentials. The forward and inverse DFT kernels are given by

$$\begin{aligned} f_{ker}(u,v;m,n) &= \tfrac{1}{N} e^{-i\frac{2\pi}{N}(um+vn)} \\ i_{ker}(u,v;m,n) &= \tfrac{1}{N} e^{+i\frac{2\pi}{N}(um+vn)} \end{aligned} \tag{2.8}$$

where $i = \sqrt{-1}$. A major feature of the DFT is the existence of a fast algorithm for its computation called the Fast Fourier transform (FFT). When N is a power of 2, the DFT requires approximately N^2 complex operations, whilst the FFT requires only $N\log_2 N$ [11, pp. 380-388].

Walsh-Hadamard Transform

The WHT [38] is also a separable unitary and symmetric transform, which has the same forward and inverse kernels. The basis kernel of the WHT is denoted by

$$f_{ker}(u,v;m,n) = i_{ker}(u,v;m,n) = \frac{1}{N}(-1)^{p(m,n,u,v)} \tag{2.9}$$

with

$$p(m,n;x,y) = \sum_{k=0}^{\log_2 N-1} (u_k m_k + v_k n_k) \tag{2.10}$$

where u_k, v_k, m_k and n_k are the bit states of the binary representations of u, v, m, and n respectively [38]. The main advantage of the WHT is that apart from the factor $\frac{1}{N}$, the

computation requires only addition and subtraction. This is in contrast to most other transforms which require multiplication as well [11, pp. 389-392]. A fast algorithm also exists that requires only $N\log_2 N$ operations. WHT is a very efficient transform for square waves in binary transmission applications. However, its energy compaction for continuous signals is not as good as that of the KLT or DFT [34].

Discrete Cosine Transform

The DCT [39] is another transform which produces an effective compaction of transform coefficients and is relatively easy to implement. The DCT has recently become the most widely used of the unitary transforms and its performance approaches the KLT [11, pp. 400-414]. The forward and inverse kernels of the DCT are given by

$$\begin{aligned} f_{ker}(u,v;m,n) &= \frac{4C(u)C(v)}{N^2} cos\left[\frac{\pi u(2m+1)}{2N}\right] cos\left[\frac{\pi v(2n+1)}{2N}\right] \\ i_{ker}(u,v;m,n) &= C(u)C(v) cos\left[\frac{\pi u(2m+1)}{2N}\right] cos\left[\frac{\pi v(2n+1)}{2N}\right] \end{aligned} \tag{2.11}$$

where

$$C(k) = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } k = 0 \\ 1 & \text{otherwise.} \end{cases}$$

The DCT can also be efficiently computed although it is not quite as efficient as the DFT or WHT [40] [41] [42].

Referring to the block diagram in Figure 2.3, the original image is divided into square blocks of $N \times N$ pixels, denoted by $f(m,n)$. Each subblock $f(m,n)$ is then cosine transformed. In standard matrix row and column notation, the general two-dimensional DCT and its inverse are given by

$$F(u,v) = \frac{4C(u)C(v)}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} f(m,n) cos\left[\frac{\pi u(2m+1)}{2N}\right] cos\left[\frac{\pi v(2n+1)}{2N}\right], \tag{2.12}$$

for $u, v = 0, 1, \cdots, N-1$, and

$$f(m,n) = C(u)C(v) \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} F(u,v) cos\left[\frac{\pi u(2m+1)}{2N}\right] cos\left[\frac{\pi v(2n+1)}{2N}\right], \tag{2.13}$$

for $m, n = 0, 1, \cdots, N-1$, respectively. Note that in DCT we use lower case to denote a spatial domain function and upper case for its transform.

By transforming the subblocks into the transform or frequency domain the majority of the statistical correlations and redundancies in most image subblocks can be removed. This is because all of the signal energy is contained in only a subset of the resulting transform coefficients [40]. By keeping only these transform coefficients, a higher compression is achieved. The reconstruction of the subblock images can be obtained by inverse transforming the decoded transform coefficients. If $\check{I}(m,n)$ is a reconstruction of the whole image, the mean square quantization error (MSE) between the original image $I(m,n)$ and the reconstructed image $\check{I}(m,n)$ can be given by

$$\text{MSE} = \frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} E\left([I(m,n) - \check{I}(m,n)]^2\right) \tag{2.14}$$

where N is, in this case, the image size. For the original image data coded using 8-bit PCM, the range of pixel values is $0 - 255$, and the image quality in terms of the peak signal to rms noise ratio [11, pp. 400-414] is given by

$$\text{PSNR} = 10log\left[\frac{255^2}{\text{MSE}}\right] \text{ dB.} \tag{2.15}$$

As an example, a simulation, carried out using the test image Lenna, is illustrated in Figure 2.5. The original image is first partitioned into square blocks of size 8×8. Each block of data is then DCT transformed as defined by equation (2.12). The resulting transform coefficients, $F(u,v)$, are then quantized using a uniform step quantization. These quantized coefficients can then be entropy coded as defined in equation (2.5) for the purposes of storage or transmission. Figure 2.5 shows the original image and reconstructed image using the DCT transform. Figure 2.5 (a) shows the original image, Figure 2.5 (b) shows the reconstructed image for a compression ratio of approximately $8 : 1$ with PSNR = 32.60 dB and Figure 2.5 (c) shows the reconstructed image for a compression ratio of approximately $16 : 1$ with PSNR = 27.25 dB. It can be seen that for high compression ratios, a blocking effect is very dominant when using the DCT transform.

(a)

(b)

(c)

Figure 2.5: Example of DCT transform image compression. (a) Original image, (b) reconstructed image for a compression ratio of 8 : 1 and (c) reconstructed image for a compression ratio of 16 : 1.

Wavelet Transform

Wavelet transforms are relatively new and originated from a group of French geophysicists working in the field of seismology [43]. The wavelet transform is basically a multiresolution signal processing technique, and is a powerful tool for the analysis of non-stationary signals such as images or fractals [44]. The wavelet transform offers excellent localisation properties in both the spatial and frequency domains and retains the details and sharp edges of an image better than the conventional Fourier-based transforms. The discrete wavelet transform decomposes the signal band into a bank of band-pass frequency subbands on a logarithmic scale, in contrast to the linear frequency bands of the Fourier transform. Since the wavelet transform is the focus of this study, more detail on the wavelet transform is therefore presented in the next chapter.

2.2.5 Pyramid Coding

The basic approach of pyramid coding is to decompose the image into various components each containing image information of a different characteristic. A block diagram illustrating pyramid coding is shown in Figure 2.6.

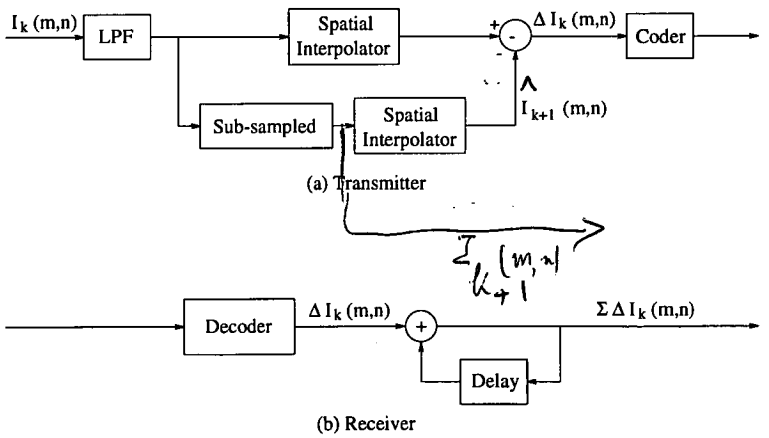
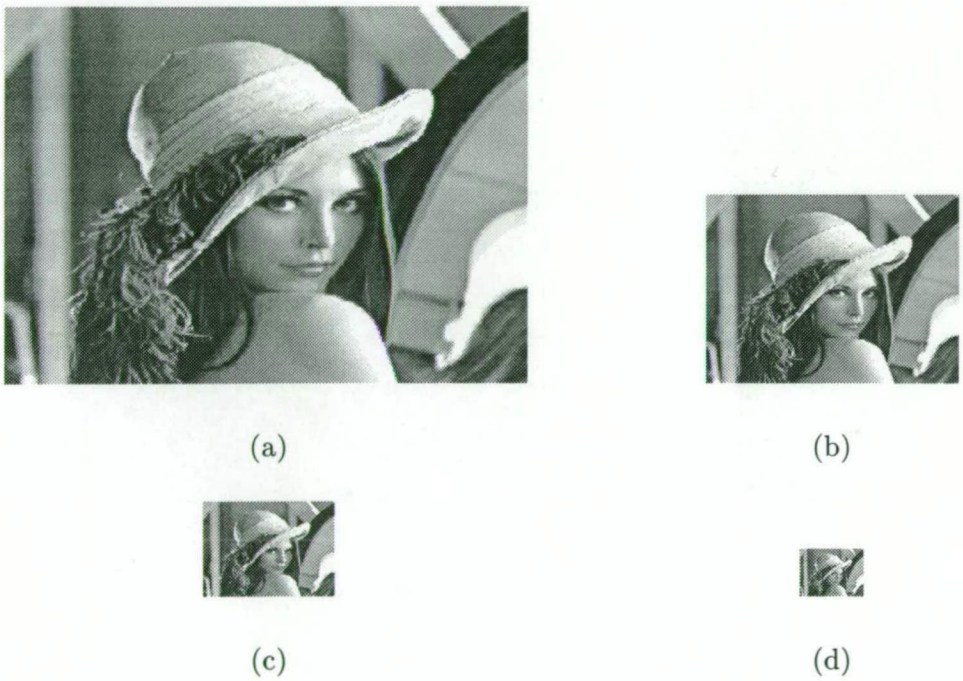


Figure 2.6: Block diagram of pyramid coding.

The input image $I_k(m,n)$, for $k = 0$, is lowpass filtered and then subsampled to generate a reduced image $I_1(m,n)$. The image $I_1(m,n)$ has a lower spatial resolution due to the lowpass filtering and a smaller size due to the subsampling than the image

$I_0(m,n)$. Spatial interpolation is used to expand each subsampled image back to the original image size. The prediction error, $\Delta I_0(m,n)$, can be obtained from the difference between the original image $I_0(m,n)$ and the output of interpolator $\hat{I}_1(m,n)$. Encoding the reduced image $I_1(m,n)$, together with prediction image $\Delta I_0(m,n)$ results in a net data compression. This is because the reduced image $I_1(m,n)$ is largely decorrelated, and can be represented pixel by pixel with many fewer bits than the original image $I_0(m,n)$. Furthermore, the reduced image $I_1(m,n)$ is lowpass filtered and can be encoded at a reduced sample rate [45]. The reduced image $I_1(m,n)$ is then lowpass filtered and subsampled to yield a second reduced image $I_2(m,n)$ and a second prediction error $\Delta I_1(m,n)$. Repeating this process generates a pyramid of prediction errors which decrease in size from bottom to top. The top image is small since it is the result of multiple lowpass filtering and subsampling. Figure 2.7 shows the original image and reduced image for a pyramid level of $k = 3$.



(a) (b)

(c) (d)

Figure 2.7: Example of pyramid coding. (a) Original image, (b) reduced image for k = 1, (c) reduced image for k = 2 and (d) reduced image for k = 3.

At the receiver, the decoding of different images $\Delta I_k(m,n)$ for $k = 0,1,...K-1$, is performed by simply adding the subsequent images as they arrive to build the reconstructed image. The pyramid subband coding depends on the type of lowpass filtering

used and how the lowpass filtered image is subsampled. Burt and Adelson [45] use a separable low-pass filter with a 5×5 point impulse response and a weighted constant factor. This pyramid coding scheme is called the Laplacian pyramid.

2.2.6 Subband Coding

Subband coding [46] [47] is similar to pyramid coding in that the original image is split up into separate frequency bands. Each image frequency band contains less correlated data and therefore less statistical redundancies than the original image. Figure 2.8 shows a block diagram illustrating subband coding. This technique splits and bandpass filters the input image to obtain the various frequency bands. Each of these subbands is then encoded separately by using either PCM or DPCM at a bit rate accurately matched to the subband statistics. Each coded subband is then time multiplexed and transmitted. At the receiver, the reconstructed image can be obtained by adding together the reconstructed bandpass images. The main advantage of subband coding is that different coding schemes can be employed in different subbands, i.e., in the higher subbands, coarse quantization may be used as opposed to the lower subbands where a finer quantization must be used. This is in contrast to pyramid coding which obtains the different frequency bands using an iterative approach in the spatial domain. This property is reminiscent of the human visual system which is described in the following section.

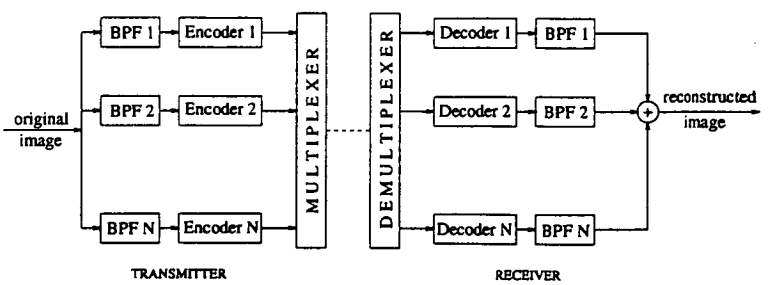


Figure 2.8: Block diagram of subband coding.

2.3 Psychovisual Image Compression Techniques

A psychovisual compression technique is one which omits any information which a human observer would be insensitive to. Information is discarded, but in a way that is not perceptible to a human observer. This technique offers much greater levels of compression, but it is no longer possible to reconstruct the original image perfectly. In practice, however, there is always a trade-off between the quality of the reconstructed image and the compression rate achieved. This trade-off can be assessed using mathematical criteria, but inevitably the best assessment is provided by a human observer. This section discusses the HVS model, which is based on the results of psychophysical and physiological experiments. The image compression techniques of the following chapter take into account the model of the HVS described in this section.

2.3.1 Visual Phenomena

Visual psychophysics refers to the study of the HVS from a transfer function point of view, where the HVS is described as an input-output system with visual stimuli as the input and the perception of these stimuli as the output. For image coding applications the main concern is with the "visibility" of artifacts or information loss introduced by various coding schemes. Two principal parameters that affect the visibility of these artifacts are introduced : contrast sensitivity and visual masking.

Contrast Sensitivity Function

It is often convenient for many psychophysical experiments to assume that the visual system is linear, although in reality the visual system is in fact non-linear. The detail of this linear assumption may be found in [48] [49] [50]. Using this linear assumption, the threshold at which arbitrarily shaped test stimuli become visible can be calculated if the transfer function or the impulse response of the linear system is known [11, pp. 256-274]. The block diagram of a simple model of such a system is shown in Figure 2.9. The input test stimulus is first linearly filtered and then if the resulting filtered output exceeds a threshold, the stimulus is considered visible. This model is successful for some stimuli, but fails for some others [11, pp. 256-274]. In order to make the system more

appropriate for these other situations, the model can be extended to a multichannel system utilizing a bank of linear filters to handle inputs of arbitrary shape, i.e. mixed frequencies.

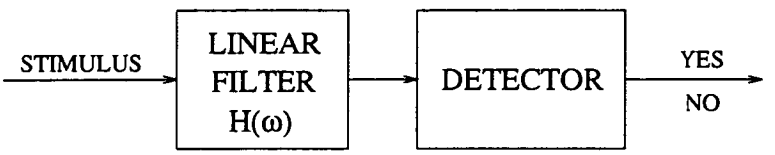


Figure 2.9: A simple single channel model of a transfer function for test stimuli.

The visibility threshold of an arbitrary shape can be derived from a transfer function or a contrast function. Campbell and Green [51] tried to measure the global transfer function of the visual system. In their psychophysical experiments, vertical sinusoidal gratings of different spatial frequencies were presented to an observer. In psychophysics, spatial frequencies are measured in cycles per degree (cpd) of visual angle subtended on the eye. The transfer function of the visual system is defined as the ratio of the contrast perceived by the observer to the real contrast of the stimulus for the sinusoidal grating. The contrast is then given by

$$C = \frac{L_{max} - L_{min}}{L_{max} + L_{min}} \tag{2.16}$$

where L_{max} and L_{min} are the maximum and minimum luminance of stimuli. In order to obtain the transfer function, a solution that is widely adopted is to measure the "contrast sensitivity function". At each frequency ω, the minimum contrast $C_{min}(\omega)$ is measured at which the sinusoidal grating becomes distinguishable from a uniform background. This contrast is called the contrast threshold. The relationship between the threshold of the contrast perception $C_{min}(\omega)$ at difference spatial frequencies and the contrast sensitivity function is defined by [51]

$$H(\omega) \propto S(\omega) = \frac{1}{C_{min}(\omega)}. \tag{2.17}$$

Many experiments have been performed to measure the visibility threshold values and used to characterize the HVS in the Fourier domain. However, it is not always feasible to measure the output for humans in a completely controlled fashion. Necessarily, psy-

chological variables are used, requiring many assumptions about the system behaviour [52]. Dooley [53] has provided the following equation to fit the experimental results of the contrast sensitivity function

$$S(\omega) = 5.05\left(1 - e^{0.1\omega}\right)\left(e^{-0.138\omega}\right). \tag{2.18}$$

Many experiments have been made, and the same alternative expression can be found with some minor variations. The general proposed mathematical form of the contrast sensitivity function $S(\omega)$ [54] is given by equation

$$S(\omega) = A(B + C\omega^D)e^{-E\omega} \tag{2.19}$$

where A, B, C, D and E are constants, and D varies from 1.0 and 2.0 and E varies from 0.13 to 0.58. Figure 2.10 shows a number of more common response curves as proposed by Kelly [55], Wilson and Giese [56], Wilson [57], King et al. [58] and Bowon et al. [59].

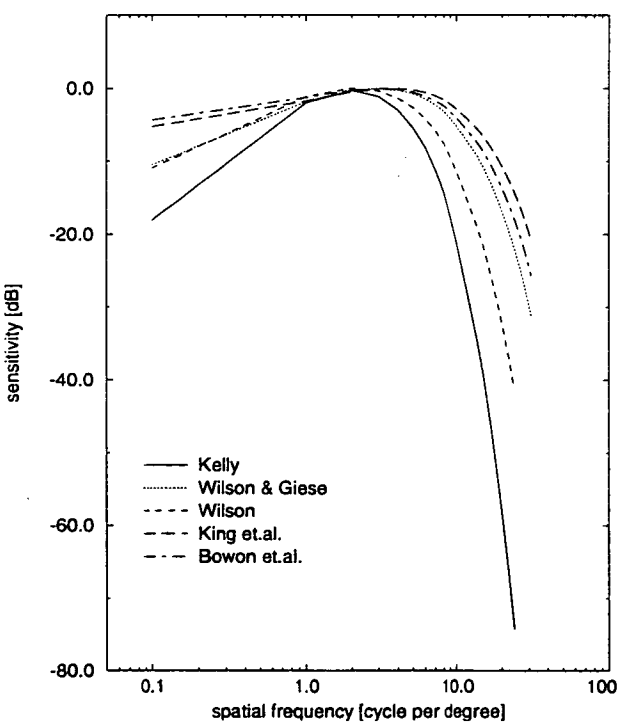


Figure 2.10: The contrast sensitivity function characteristic which has been explored by Kelly [55], Wilson and Giese [56], Wilson [57], King et al. [58] and Bowon et al. [59].

The response curves show that the contrast sensitivity function has a bandpass frequency response with a peak sensitivity between 3 to 3.5 cycles per degree (cpd). This sensitivity is reduced slightly at low frequencies less than 1 cpd, but drops off very sharply at higher frequencies of more than 10 cpd. Therefore, if an image has a spatial frequency range lower than 0.1 cpd and higher than 10 cpd, then these portions of the frequency spectrum can be discarded with minimum degradation in the reconstructed image.

Visual Masking

The term of masking is used to refer to the action of a visual stimulus on the visibility of another [60]. The effect of masking implies that noise in regions of high activity is less perceptible than those in the low activity regions [61]. Visual masking effects can be exploited in image coding by allowing a greater quantization error in regions of high activity. Figure 2.11 shows the results of a very simple experiment [62] in which a thin vertical line of light is moved across a dark-light transition or edge. As an arbitrary stimulus is moved across this dark-light transition, the visibility threshold ΔL increases near the boundary and then decreases as the stimulus is moved further away. These results correspond to a reduced visibility threshold for image distortion in areas of high detail or high spatial activity. If a suitable measure of the amount of activity in different regions of an image can be defined, then the masking effect can be used to allow for more distortion in areas where activity is high [54].

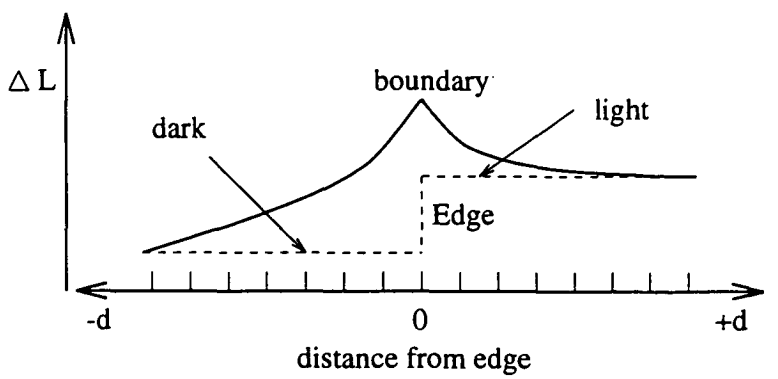


Figure 2.11: The results of a simple experiment showing the relationship between visibility threshold and the distance from a dark-light transition.

Although a number of methods to measure the amount of activity in images have been proposed in the past [11, pp. 256-274] [60] [61] [63] [64] [65], the method using a weighted summation of vertical and horizontal intensity gradients in the spatial domain, as well as energy values in the frequency domain, is the most widely accepted [65]. By using masking effects, the efficiency of coding schemes can be increased further improving the compression process [60] [66] [67].

2.4 Combining Compression Techniques

As mentioned in Sections 2.2 and 2.3 statistical and psychovisual compression techniques simply exploit the statistical and psychovisual redundancies respectively. In order to produce a higher compression ratio for visual images, both statistical and psychovisual redundancies must be removed from an image in the coding process.

Statistical compression techniques (PCM, DPCM, transform coding, subband coding, pyramid coding) all use a form of amplitude quantization in their algorithms to improve compression performance. Simple quantization alone, however, is not the most efficient or flexible techniques to combine with a statistical compression algorithm [54]. The combination of quantization and psychophysics, on the other hand has the potential to remove most subjectively redundant information efficiently from an image, a process which is based on the actual behaviour of the HVS [54]. Furthermore, subband coding and pyramid coding schemes can be combined with visual psychophysics-based compression techniques, since both of these statistical schemes break the original image data down into separate frequency bands. A process that is similar to the bandpass filter characteristics of the HVS and can be used to quantize the information in each band depending on the relative frequency band. These two coding schemes, however, are unable to obtain the high compression ratios required for broadband Integrated Services Digital Networks (ISDN) image-based applications [54].

Transform coding is able to achieve optimum statistical compression ratios, especially the DCT transform. Much research has been performed in combining the DCT transform coding and visual psychophysics-based compression techniques [58] [59] [67] [68] resulting in a higher compression ratio and good reconstruction of the original image. Unfortunately, these schemes produce the blocking effect noticed in Figure 2.5(c) at low

bit rates. Wavelet transform coding can effectively eliminate this blocking effect [18] [69] [70] because the wavelet bases in adjacent subbands overlap one another. Another advantage of wavelet transform coding is that it is very similar to subband coding [71] [72].

A wavelet transform coding scheme which incorporates psychovisual compression techniques is outlined. The block diagram of this scheme is illustrated in Figure 2.12. The wavelet transform is basically a multiresolution decomposition similar in form to the psychophysical and psychological models of the HVS, a property noticed by Mallat [69], Wilson [73] and Nguyen [74]. Therefore, for image coding purposes, multi-band image decompositions can be matched to the HVS sensitivity function and used to take advantage of the intrinsic statistical properties of images. The sensitivity function illustrated in Figure 2.10 shows that the sensitivity of the HVS depends upon the frequency of the stimulus. If we use this sensitivity function in conjunction with an image decomposed using the wavelet transform, we can then discard some transform coefficients or use a coarser quantization resulting in a potential reduction in the number of bits per pixel and an improvement in the compression ratio.

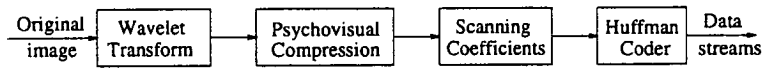


Figure 2.12: Block diagram of psychovisual compression using the wavelet transform.

2.5 Video Coding

All the compression techniques mentioned above, which involve spatial correlations in single frames where the redundancies are exploited either statistically or subjectively, are known as intraframe coding techniques. Interframe coding techniques, by contrast, attempt to exploit the redundancies produced by temporal correlations as well as spatial correlations in successive video signals. These techniques hold the promise of a significantly greater reduction in the data required to transmit the video signal as compared to intraframe coding.

The simplest interframe coding technique is called "conditional replenishment" [75] [76] [77]. This technique bases the coding scheme on the previous frame and is also often called predictive coding. In the conditional replenishment technique, only pixels the values of which have changed significantly since the last frame, as compared to a certain threshold, are transmitted. Another technique, which still uses predictive coding from the previous frame, is adaptive intra-inter-frame prediction [78]. In this technique, interframe prediction is used for scenes in images where there is little motion, while intraframe prediction is used for areas where this is much motion. The switching between intra- and inter-frame prediction or a combination of both, is usually controlled by the signal changes of previously transmitted pixels so that no overhead control information needs to be sent [21]. The prediction error can be quantized and transmitted for every pixel or can be thresholded into predictable and non-predictable pixels [79].

Adaptive prediction displacement of a moving object which is based on information obtained from successive frames is known as "Motion Compensation". This scheme was studied by Limb and Murphy [80], and Cafforio and Rocca [81] by measuring small displacements based on a very simple model of moving objects in a stationary background scene for segmentation purposes. A later refinement developed by Netravali et al. [82], Cafforo et al. [83] and Robbin et al. [84] led to one set of techniques known as "Pel Recursive Motion Compensation", which recursively adjusts the translational estimates at every pixel or every small block of pixels. Jain and Jain [85] developed another technique known as "Block Matching Motion Estimation". This technique estimates the location of a block of pixels in the current frame by using a search in a confined window defined in the previous frame. Location of the block results in the displacement vector for that block. Different search methods have been proposed to avoid an exhaustive search [86] [87] [88], [89].

In order to produce a higher compression ratio, transform coding has been applied to video coding, and can be carried out as a three-dimensional transform [90] or in an interframe coding scheme [85] [91]. In the latter case motion compensation can be performed in either the spatial domain or the frequency domain. Transform coding can also be combined with predictive coding so that the transform coefficients from intraframe transformations of the previous frame can be used to predict the transform coefficients of the current frame [92]. The latest CCITT H.261 Recommendations [36], the JPEG standard [6] and the MPEG draft [93], are also DCT transform based and are intra-inter-frame adaptive with optional motion compensation. Their schemes,

however, result in a blocking effect for low bit-rates. As mentioned in Section 2.4 the blocking effect, can be reduced by replacing the DCT transform with the wavelet transform. Furthermore the wavelet transform and HVS can be combined resulting in a coding scheme that does not produce the blocking effect and results in a higher compression ratio. This scheme will be presented in the last chapter of this thesis.

2.6 Summary

Visual images of interest contain a considerable amount of statistically and subjectively redundant information. Statistical compression techniques remove only statistical redundancies from an image by exploiting the statistical correlations and predictability of the image data with little or no information being lost. Therefore, the reconstructed image produced from these techniques can be almost perfectly recovered. Psychovisual compression techniques are irreversible compression techniques. These techniques discard the subjective redundancies in an image and hence change the original image. As long as only subjective redundant information is removed from the image the difference between the original and the reconstructed image will not be noticed by a human observer.

There is no difference between image coding and video coding if video image sequences have no correlation between successive video frames. Video coding can then be treated as image coding, resulting in the same overall performance. Typical image sequences, however, have correlations between successive images in the sequence. By exploiting the temporal correlations between successive images a larger compression ratio can be obtained. There are many techniques which have been described above, that exploit the redundancies in the images either in single frames or in sequences of images. The combination of transform coding, adaptive intra-inter-frame motion compensation and the taking into account of the HVS system results in a low bit-rate, and will be of benefit for video coding applications.

Chapter 3

Wavelet Transforms

3.1 Introduction

Despite being a relatively new area of mathematics, wavelet theory has actually been known since the beginning of this century [94] [95]. It has, however, only recently been developed as a unified framework in the mid-eighties by a group of French geophysicists, theoretical physicists and mathematicians, namely Morlet, Grossman, Meyer and Lemarie; and in the late eighties by a French signal processing group, namely Daubechies and Mallat. Wavelet theory represents signals by breaking them down into scaled and translated wavelets [96]. The breaking down process is termed a wavelet transform or wavelet decomposition. Wavelet reconstruction or the inverse wavelet transform involves a weighted summation of the scaled and translated wavelets so that the original signal is retrieved.

The wavelet transform is basically a multiresolution or multi-scale approach to signal analysis [18] [97]. In various signal processing applications, the wavelet transform has become a useful tool that transforms a signal into an alternative domain, i.e. the wavelet domain, so that certain features of the signal can be more easily identified. The wavelet transform has an important advantage over the Fourier transform, in which the basis functions are complex sinusoids. The basis functions of the Fourier transform are perfectly localised in frequency but global in space or time. Thus the

Fourier transform domain cannot show the location of discontinuities in a signal. In contrast, in the wavelet transform the basis functions are localised in both the time and frequency domains. These basis functions are obtained from a mother wavelet ψ by dilation and translation. The dilation and translation properties allow the wavelet bases to analyze the signal at different scales (frequencies) and locations.

This chapter introduces some wavelet bases and details their design based on the work of Daubechies [98] [99] [100]. All these wavelet bases will be applied in the study presented in the next chapter.

3.2 Continuous Wavelet Transform

The Wavelet Transform (WT) of a given signal may be interpreted as the decomposition of the signal into a set of time-frequency functions by the use of translated and dilated basis functions of a mother wavelet. In other words the WT breaks a signal into two variables: frequency or scale and time.

The wavelet basis functions are obtained from a single mother wavelet by translation and dilation [18] [101] [102] [103], i.e.

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right) \tag{3.1}$$

where a is the scale or dilation parameter, b is the translation parameter and a, b vary continuously over $\mathbf{R}$ and $a \neq 0$. For large a, the basis function become stretched (dilated) versions of the mother wavelet, and can be used to analyze low frequency components. For small a the basis functions become contracted, and can be used for high frequency components. The mother wavelet function must be an admissible function, that satisfies the following equation [101],

$$C_\psi = \int_{-\infty}^{\infty} \frac{\left|\hat{\psi}(\omega)\right|^2}{|\omega|} < \infty \tag{3.2}$$

where C_ψ is a constant. The mother wavelet function can be normalized further to have an admissibility constant, that is unity. This means that $\psi(x)$ is necessarily the impulse response of a bandpass filter. In the time domain, this results in a zero value

for the integral of the mother wavelet, i.e.

$$\int \psi(x)dx = 0. \tag{3.3}$$

Therefore, the Continuous WT (CWT) of a signal f is defined as

$$WTf(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \psi^*(\frac{x-b}{a}) f(x)dx, \tag{3.4}$$

where the superscript $*$ denotes the complex conjugate. This equation is discussed in more detail in [101] [104]. The time-frequency resolution of the wavelet transform implies that at high frequencies the wavelet transform is smaller in time resolution, while at low frequencies the wavelet transform is wider in time resolution. Figure 3.1 shows the basis functions and time-frequency resolution of the wavelet transform.

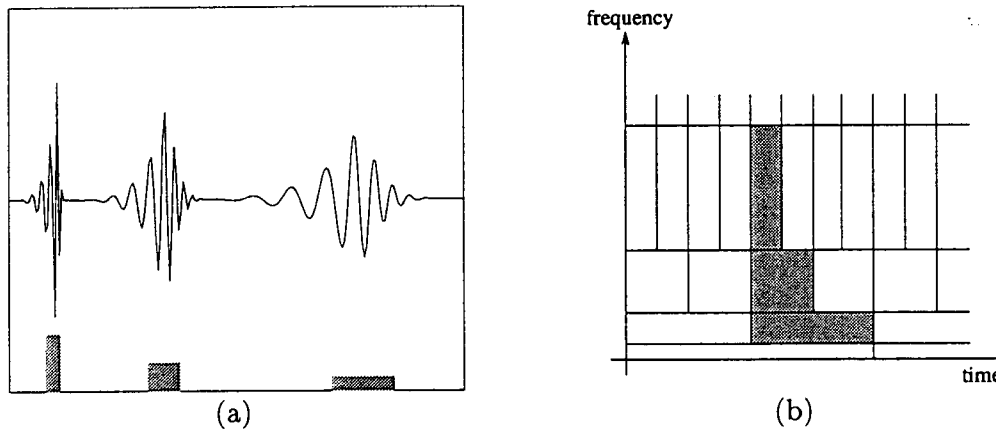


Figure 3.1: Typical basis functions and the time-frequency resolution of the wavelet transform, (a) basis functions and (b) the time-frequency plane.

3.3 Discrete Wavelet Transform

The wavelet transform described in equation (3.4) is highly redundant when the dilation a and translation b are continuous. In order to reduce this redundancy a discrete version of the wavelet transform can be defined where both a and b take only discrete values.

Equation (3.1) can therefore be rewritten as

$$\begin{aligned} \psi_{j,n}(x) &= a_0^{-j/2}\psi(a_0^{-j}(x - nb_0a_0^j)) \\ &= a_0^{-j/2}\psi(a_0^{-j}x - nb_0) \end{aligned} \tag{3.5}$$

where a corresponds to a_0^j and $b = na_0^jb_0$. $j, n \in \mathbf{Z}$, $a_0 > 1$, $b_0 \neq 0$ and ψ is admissible. The discrete wavelet transform is then defined as

$$WTf(j,n) = a_0^{-j/2}\int_{-\infty}^{\infty}\psi^*(a_0^{-j}x - nb_0)f(x)dx. \tag{3.6}$$

Figure 3.2 shows schematically the lattice of time-frequency localization centers corresponding to the basis functions $\psi_{j,n}$.

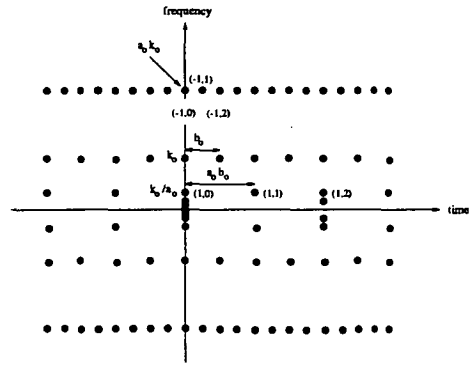


Figure 3.2: The lattice of time-frequency location centers corresponding to $\psi_{j,n}$.

Of particular interest is the discretization onto a dyadic grid, where $a_0 = 2$ and $b_0 = 1$. It is possible to construct a function ψ, so that the set of scaled and translated functions constitute an orthonormal set of basis functions,

$$\psi_{j,n}(x) = 2^{-j/2}\psi(2^{-j}x - n) \quad j, n \in \mathbf{Z}. \tag{3.7}$$

This orthonormality gives,

$$\langle\psi_{j,n}(x), \psi_{k,l}(x)\rangle = \delta_{j-k}\delta_{n-l}. \tag{3.8}$$

A simple example of a wavelet whose basis functions constitute an orthonormal set is the Haar wavelet. The Haar wavelet is defined by

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \leq x < \frac{1}{2} \\ -1 & \text{if } \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

The orthonormality is easily verified since at a given scale, the translates are non-overlapping, and because of the scale change of 2, the basis function are orthonormal across scale. Figure 3.3 shows the Haar wavelet bases.

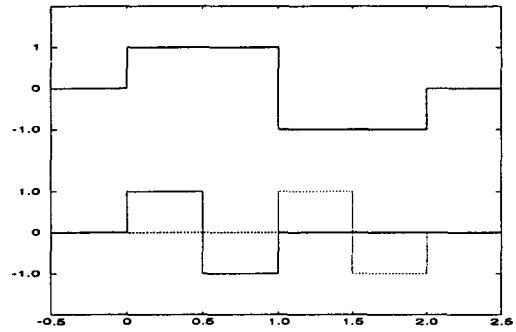


Figure 3.3: The orthogonal system of scaled and translated wavelets of the Haar basis. The upper plot shows $\psi(\frac{x}{2})$ of the dilated version and the lower shows the translated versions of $\psi(x)$ and $\psi(x-1)$.

3.4 Multiresolution Analysis

The principle of multiresolution analysis can be interpreted by writing a $\mathbf{L}^2(\mathbf{R})$ function f as a limit of successive approximations, each of which is a smoothed version of f [18] [69] [100] [105] [106]. A ladder of lowpass spaces $(V_j)_{j\in\mathbf{Z}}$ is introduced to represent the successive resolution levels such that $V_j = \{f \in \mathbf{L}^2(\mathbf{R}); \forall k \in \mathbf{Z} : f|_{[2^j k, 2^j(k+1)[} = \text{constant}\}$. These spaces have the following properties:

1. $\cdots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \cdots$
2. $\bigcup_{j\in\mathbf{Z}} V_j = \mathbf{L}^2(\mathbf{R})$ and $\bigcap_{j\in\mathbf{Z}} V_j = 0$
3. $f(\cdot) \in V_j \longleftrightarrow f(2\cdot) \in V_{j-1}$

4. There exists a unique function $\phi \in V_0$, called a lowpass scaling function, such that the $\phi_{j,n}$ for all $j \in \mathbf{Z}$, constitute an unconditional for V_j, that is $V_j =$ linearspan$(\phi_{\mathrm{j,n}}), \mathrm{n} \in \mathrm{Z}$

Property 3 implies that if $f(\cdot) \in V_j$ then the translates $f(\cdot - 2^j n) \in V_j$ for $n \in \mathbf{Z}$ and there exists $\phi \in V_0$ so that $\phi_{0,n}$ for $n \in \mathbf{Z}$ is an orthonormal basis in V_0 for $j, n \in \mathbf{Z}$, such that $\phi_{j,n}(x) = 2^{-j/2}\phi(2^{-j}x - n)$. This implies further that $\phi_{j,n}$ is an orthonormal basis for V_j for all $j \in \mathbf{Z}$, and

$$Proj(_{V_j})f = \sum_{n \in \mathbf{Z}} \langle f, \phi_{j,n} \rangle \phi_{j,n}. \tag{3.10}$$

Since $\phi \in V_0 \subset V_{-1} = span\{\phi_{-1,n}; n \in \mathbf{Z}\}$, $\phi(2x - n)$ may be used as a basis for $\phi(x)$. The function ϕ therefore necessarily satisfies an equation of the type [99] [100],

$$\phi(x) = 2^{1/2} \sum_{n=-\infty}^{\infty} h(n)\phi(2x - n) \tag{3.11}$$

where infinitely many of the weighted $h(n)$ may differ from zero. Equation (3.11) implies that a scaling function can be obtained from a linear combination of its scaled versions. Also since $\langle \phi(x), \phi(x - k) \rangle = \delta_k$ we obtain

$$\sum_n h(n)h(n - 2k) = \delta_k. \tag{3.12}$$

The sequence $h(n)$ is orthonormal with respect to even shifts.

Next, let W_j be an orthogonal bandpass complement of V_j in V_{j-1}. That is, $V_{j-1} = V_j \oplus W_j$. In other words, V_{j-1} is equivalent to V_j plus some added detail corresponding to W_j. Since $W_j \subset V_{j-1}$, it clear that the basis functions of W_j can also be written in terms of these V_{j-1}, i.e.

$$\psi(x) = 2^{1/2} \sum_{n=-\infty}^{\infty} g(n)\phi(2x - n). \tag{3.13}$$

Due to W_j and V_j being orthogonal, then

$$\begin{aligned}\langle\phi(x),\psi(x-k)\rangle &= 2\langle\sum_n h(n)\phi(2x-n),\sum_l g(l)\phi(2x-2k-2l)\rangle \\ &= 2\sum_n\sum_l h(n)g(l)\tfrac{\delta_{l+2k-n}}{2} \\ &= \sum_n h(n)g(n-2k)=0.\end{aligned} \tag{3.14}$$

Since $\phi(x)$ is orthonormal then W_j is orthonormal and therefore $\langle\psi(x),\psi(x-k)\rangle=\delta_k$, which gives,

$$\sum_n g(n)g(n-2k)=\delta_k. \tag{3.15}$$

Therefore, the $\psi_{j,n}(x)=2^{-j/2}\psi(2^{-j}x-n)$ for $j,n\in\mathbf{Z}$ constitute an orthonormal basis for $L^2(\mathbf{R})$. Moreover, for every fixed j, the $\langle f,\psi_{j,n}(x)\rangle$ expresses the difference between the approximations of f with resolution 2^j and 2^{j-1}, i.e.,

$$Proj(V_{j-1})f = Proj(V_j)f + \sum_{n\in Z}\langle f,\psi_{j,n}\rangle\,\psi_{j,n}. \tag{3.16}$$

Figure 3.4 shows the octave structure of the spaces, where V_j contains the frequencies $(0,2^{-j}\pi)$. The V_j are successive ideal lowpass spaces. The projection of a signal $f(x)$ onto V_j will thus give a lowpass approximation of the function. W_j contains the frequencies $(2^{-(j+1)}\pi,2^{-j}\pi)$.

3.5 Subband Filtering and Multiresolution Analysis

In multiresolution analysis, the wavelet coefficients c_n^j and d_n^j can be derived from equations (3.11) and (3.13), i.e.,

$$\begin{aligned}c_n^j &= \sum_k h_{k-2n}c_k^{j-1} = h(-k)\odot c^{j-1}(k) \\ d_n^j &= \sum_k g_{k+1-2n}c_k^{j-1} = g(-k-1)\odot c^{j-1}(k).\end{aligned} \tag{3.17}$$

Both the formulae in equation (3.17) have the structure of a convolution, followed by a down-sampling of 2, i.e. only one out of every two entries of the convolution is retained. In other words, the decomposition of a signal can be computed by a coarser

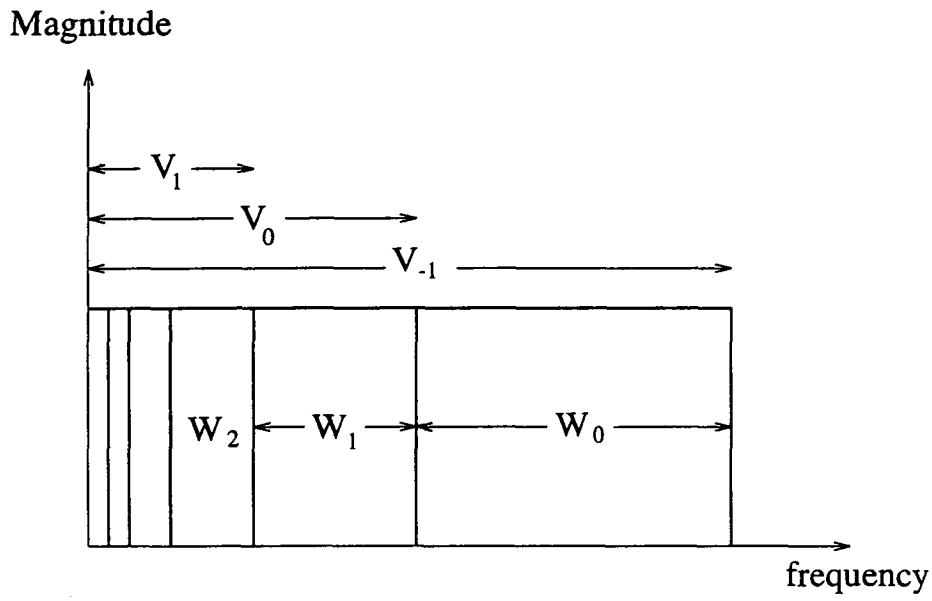


Figure 3.4: The octave structure of V_j and W_j.

approximation from a finer one, as well as from the difference in information between the two successive approximations. From equations (3.10) and (3.16), on the other hand, we obtain in the reverse direction,

$$c_l^{j-1} = \sum_n \left[\bar{h}_{l-2n} c_n^j + g_{l-2n} d_n^j \right]. \tag{3.18}$$

The right hand side of equation (3.18) can be viewed as a successive up-sampling by two of the coefficients c_n^j and d_n^j, where the coefficients are considered as the even entries of a sequence whose odd entries are zeros. These two up-sampled sequences are then convolved with the filter coefficients $\bar{h}(n)$ and $\bar{g}(n)$ respectively, and the two are then added to result in the original signal. This process is called signal reconstruction. In signal processing, these decomposition or analysis and reconstruction or synthesis schemes are known as subband filtering with exact reconstruction. Figure 3.5 shows the block diagram of the subband filtering with exact reconstruction.

In order to avoid the awkwardness of dealing with the mirror-image filters $h(-n)$ and $g(-n)$ we rewrite the decomposition or analysis equations above to have proper convo-

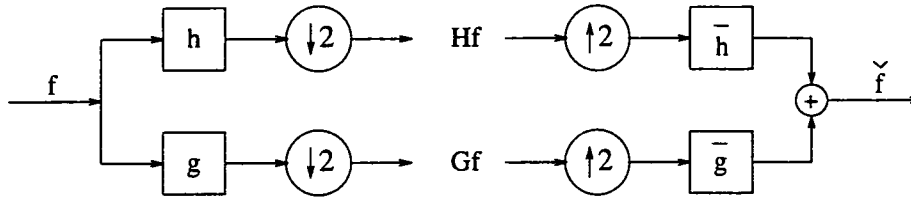


Figure 3.5: Subband filtering scheme with exact reconstruction.

lution

$$\begin{aligned} c_n^j &= \sum_k h_{2n-k} c_k^{j-1} \\ \text{and} \quad d_n^j &= \sum_k g_{2n-k+1} c_k^{j-1}. \end{aligned} \tag{3.19}$$

Similarly, the synthesis equation is rewritten as

$$c_l^{j-1} = \sum_n \left[\bar{h}_{2n-l} c_n^j + \bar{g}_{2n-l} d_n^j \right]. \tag{3.20}$$

By inserting equation (3.19) for c_n^j and d_n^j respectively into equation (3.20), it is obviously that for exact reconstruction

$$\sum_n \left[h_{2n-l} \bar{h}_{2n-k} + h_{2n-l+1} \bar{h}_{2n-k+1} \right] = \delta_{lk}. \tag{3.21}$$

We next carry out the analysis in the z-domain. Let $F(z) = c_0 + c_1 Z + c_2 Z + c_3 Z + \cdots$. The transform of the sequence formed by retaining every second sample (i.e. only even power of z) is $\frac{1}{2}(F(z) + F(-z))$. Equations (3.19) and (3.20) are therefore

$$\begin{aligned} c^j(z^2) &= \tfrac{1}{2} \left[H(z) c^{j-1}(z) + H(-z) c^{j-1}(-z) \right] \\ z d^j(z^2) &= \tfrac{1}{2} \left[G(z) c^{j-1}(z) - G(-z) c^{j-1}(-z) \right] \\ c^{j-1}(z) &= \bar{\bar{H}}(z) C^j(z^2) + \bar{\bar{G}}(z) z d^j(z^2) \end{aligned} \tag{3.22}$$

where $H(-z) = \sum (-1)^n h_n z^n$ and $\bar{H} = \sum h_{-n} z^n = \sum h_n z^{-n} = H(z^{-1})$ is the mirror-image of h_n. For $|z| = 1$, $\bar{H} = H^*(\omega)$ if a_n is a real sequence. By using (3.22) and by equating the $c^{j-1}(z)$ and the $c^{j-1}(-z)$ terms separately, we obtain

$$\frac{1}{2} \left[\bar{\bar{H}}(z) H(z) + \bar{\bar{G}}(z) G(z) \right] = 1 \tag{3.23}$$

and

$$\frac{1}{2}\left[\overline{\bar{H}}(z)H(-z) - \overline{\bar{G}}(z)G(-z)\right] = 0. \tag{3.24}$$

In equation (3.23), $H(z)$ and $G(z)$ cannot have common zeros, therefore $H(-z)$ and $G(-z)$ in equation (3.24) cannot have common zeros. That is $\overline{\bar{H}}(z) = 0$ whenever $G(-z) = 0$ and $\overline{\bar{G}}(z) = 0$ whenever $H(-z) = 0$. Therefore

$$\overline{\bar{H}}(z) = p(z)G(-z),$$

and

$$\overline{\bar{G}}(z) = p(z)H(-z),$$

where $p(z)$ does not have its own zeros. Inserting these into equation (3.23), we have

$$p(z)\left[H(z)G(-z) + H(-z)G(z)\right] = 2.$$

Let the only possible $p(z)$ be

$$p(z) = \alpha z^k, \alpha \neq 0$$

then

$$\overline{\bar{H}}(z) = \alpha z^k G(-z)$$

and

$$\overline{\bar{G}}(z) = \alpha z^k H(-z).$$

Since $\overline{z^{+k}F(z)} = z^{-k}\overline{F}(z)$, the inverse transforms give

$$\begin{aligned} g_n &= \alpha^{-1}(-1)^n \bar{h}_{-n-k} \\ \bar{g}_n &= \alpha(-1)^{n+k} h_{-n-k}. \end{aligned} \tag{3.25}$$

Equations (3.23) and (3.24) may now be rewritten as

$$\frac{1}{2}\left[\overline{\bar{H}}(z)H(z) + \overline{\bar{H}}(-z)H(-z)\right] = 1,$$

or its inverse convolution, retaining only every second samples, is

$$\sum \bar{h}_n h_{n+2k} = \delta_{k0} \tag{3.26}$$

where k allows us to shift to a convenient origin for $g(n)$ in relation to $\bar{h}(n)$. If $g(n)$ and $\bar{h}(n)$ are both to start from $n \geq 0$ for real time processing, then

$$g(n) = (-1)^n \bar{h}(\bar{N} - 1 - n)$$
$$\bar{g}(n) = (-1)^n h(N - 1 - n)$$

where $\bar{N}$ and N are the lengths of $\bar{h}(n)$ and $h(n)$ respectively. For the orthonormal case where $h \equiv \bar{h}$ and $g \equiv \bar{g}$, therefore

$$g(n) = (-1)^n h(N - 1 - n)$$

where N has to be even.

For multiresolution analysis, any sequence in the original space $l^2(\mathbf{Z})$ which is denoted by V_{j-1} can be reconstructed as the sum of its projections on two orthogonal subspaces $V_{j-1} = V_j \oplus W_j$. In other words, $l^2(\mathbf{Z})$ can be written as the direct sum of the row spaces of $\mathbf{H}$ and $\mathbf{G}$ [72]. By interleaving rows $\mathbf{H}$ and $\mathbf{G}$, we can obtain a matrix block Toeplitz orthonormal $\mathbf{A}$,

$$\mathbf{A} = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ h(N-1) & h(N-2) & \cdots & \cdots & h(0) & 0 & 0 & \\ g(N-1) & g(N-2) & \cdots & \cdots & g(0) & 0 & 0 & \\ \ddots & & & & & & & \ddots \\ 0 & 0 & h(N-1) & \cdots & h(2) & h(1) & h(0) & \\ 0 & 0 & g(N-1) & \cdots & g(2) & g(1) & g(0) & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix} \tag{3.27}$$

which satisfies

$$\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I}. \tag{3.28}$$

That is, the two filter impulse responses $h(n)$ and $g(n)$, together with their even translates, form an orthonormal basis for $l^2(\mathbf{Z})$. A filter pair $h(n)$ and $g(n)$, which satisfy this condition are known as quadrature mirror filters (QMF).

3.6 Design of Compactly Supported Orthonormal Wavelet Bases

Wavelets are characterised by a number of parameters, often mutually exclusive. These are spatial compactness (i.e. spatial support of wavelet basis), orthonormality, regularity or smoothness, and symmetry or anti-symmetry. The construction of orthonormal symmetrical or anti-symmetrical wavelets which are also real and compactly supported is not possible, with the exception of the Haar wavelet which is asymmetric [99]. Biorthonormal wavelet bases, however, in which the analysis (decomposition) and synthesis (reconstruction) filters are different can be symmetrical or anti-symmetrical [100]. They can also be designed so that they are compactly supported. However, the disparity between lengths and coefficients of the analysis and synthesis filters results in a more computationally expensive implementation. The disparity in the frequency responses of the two filters also causes complications in the frequency coding, as will be shown in the subjective quantization and thresholding of wavelet coefficients for incorporation into the human visual system [1] [2]. In general, orthonormal compactly supported wavelet bases are obtained from the iteration of a dilation equation having a finite number of coefficients, and therefore the wavelets can be at most piecewise smooth and cannot be expressed in a closed analytical form. A typical example is the Daubechies wavelet [99].

In many applications, symmetry and compactness are the two most desirable features of wavelets [99] [107] [108]. Compactness is obviously desirable for efficient computation and accurately detecting localised irregularities. Unfortunately, compactness and smoothness are mutually exclusive for orthonormal wavelets. The trade-off between spatial and frequency characteristics prevails in all transforms. Spatial compactness requires a broad frequency spectrum. Compactly supported wavelets therefore have a high frequency content and cannot be smooth. In image coding applications, quantization errors around edges in the images are more irritating to human eyes than symetrical errors. Asymmetrical filters, therefore require finer quantization, resulting in less image compression [99, pp. 254]. Furthermore, symmetrical filters produce less border effects. The use of smooth wavelets may minimise the quantization errors in images but smooth wavelets can only be realized by using long filter coefficients. In image coding, however, short filters are desirable in order to decrease computation time. These requirements preclude many other desirable properties and narrows down the

choice of wavelets.

3.6.1 Maximum Vanishing Moment Orthonormal Wavelet Bases

Wavelet bases associated with multiresolution analysis are characterized by the dilation equations denoted by

$$\begin{aligned}\phi(x) &= \sqrt{2}\sum_n h(n)\phi(2x-n)\\ \psi(x) &= \sqrt{2}\sum_n g(n)\phi(2x-n).\end{aligned} \tag{3.29}$$

Daubechies [98] has pointed out that to construct orthonormal wavelet bases in which ψ is compactly supported, it is necessary to choose a scaling function ϕ with compact support. The filter must then have some regularity constraints. This condition is given by $\sum_n h(n) = \sqrt{2}$. This leads to the definition of the filter H,

$$H(\omega) = \frac{1}{\sqrt{2}}\sum_n h(n)e^{-in\omega} \tag{3.30}$$

where $\{h(n)\}, n = 0, ...N-1$ are the coefficients of filter H. The function $H(\omega)$ is a 2π-periodic trigonometric polynomial for compactly supported wavelets, and it should satisfy the equation

$$|H(\omega)|^2 + |H(\omega+\pi)|^2 = 1. \tag{3.31}$$

From the subband filtering scheme, it can be shown that exact reconstruction implies that

$$|G(\omega)|^2 + |H(\omega)|^2 = 1 \tag{3.32}$$

where $G(\omega)$ and $H(\omega)$ form a QMF pair. From the orthonormality constraint the relationship between $G(\omega)$ and $H(\omega)$ can be denoted by

$$G(\omega) = e^{-i\omega}\overline{H(\omega+\pi)} \tag{3.33}$$

or equivalently by

$$g(n) = (-1)^n h(N-1-n) \tag{3.34}$$

where N is filter length and $\int \phi(x)dx = 1$.

Equation (3.29) can be equivalently expressed in the Fourier transform domain by

$$\begin{aligned} \hat{\phi}(\omega) &= H(\tfrac{\omega}{2})\hat{\phi}(\tfrac{\omega}{2}) \\ \hat{\psi}(\omega) &= G(\tfrac{\omega}{2})\hat{\phi}(\tfrac{\omega}{2}). \end{aligned} \tag{3.35}$$

From the substitution of equation (3.35) into equation (3.32) it can be readily seen that

$$\left|\hat{\phi}(\omega)\right|^2 + \left|\hat{\psi}(\omega)\right|^2 = \left|\hat{\phi}(\frac{\omega}{2})\right|^2 \tag{3.36}$$

which implies for $\omega \neq 0$ that,

$$\left|\hat{\phi}(\omega)\right|^2 = \sum_{j=1}^{\infty} \left|\hat{\psi}(2^j\omega)\right|^2. \tag{3.37}$$

The limit $j = \infty$ ensures that $\hat{\phi}(\omega)$ covers the entire positive frequency axis including $\omega = 0$.

For compactness, only a finite number of coefficients $h(n)$ and $g(n)$ of the filters can be non-zero. Therefore $H(\omega)$ and $G(\omega)$ have to be trigonometric polynomials [99]. Another parameter for a compactly supported wavelet is the regularity or smoothness of the wavelet. For an iterated function to converge to a continuous or even a piecewise continuous limit, the filter $H(\omega)$ should have a sufficient number of zeros at $\omega = \pi$ to attenuate repeat spectra in equation (3.31). For L multiple zeros at $\omega = \pi$, $H(\omega)$ can then be factorized as [99]

$$H(\omega) = \left(\frac{1 + e^{-i\omega}}{2}\right)^L Z_L(\omega) \tag{3.38}$$

where $Z_L(\omega)$ is a 2π-periodic trigonometric polynomial in C^{L-1} and $L \geq 1$. Equation (3.38) can be rewritten as

$$|H(\omega)|^2 = \left(cos^2\frac{\omega}{2}\right)^L \mathcal{L}(\omega) \tag{3.39}$$

where $Z_L(\omega) = |\mathcal{L}(\omega)|^2$ is polynomial in $cos(\omega)$ [99] such that

$$\mathcal{L}(\omega) = P(sin^2\frac{\omega}{2}).$$

Using the requirement for exact reconstruction and orthonormality, as expressed in

equation (3.31), equation (3.39) can be rewritten in terms of P as

$$(1-y)^L P(y) + y^L P(1-y) = 1 \tag{3.40}$$

where $y = sin^2(\frac{\omega}{2})$. Equation (3.40) has the form of Bezout's equation and the solution is [99]

$$P(y) = \sum_{k=0}^{L-1} \begin{pmatrix} L-1+k \\ k \end{pmatrix} y^k + y^L \mathcal{R}(\frac{1}{2} - y) \tag{3.41}$$

where $\mathcal{R}$ is an odd polynomial of $cos(\omega)$ chosen such that $P(y) \geq 0$ for $y \in [0,1]$.

For different choices of $\mathcal{R}$ and the multiplicity L of zeros at $\omega = \pi$, different compactly supported orthonormal wavelets result from the spectral factorization of equation (3.39) to obtain $H(\omega)$. The choice of $\mathcal{R}$ equal to 0 results in the construction of a family of wavelet bases called Daubechies wavelets. Therefore, $\mathcal{L}(\omega)$ can be factorized from $\mathcal{L}(\omega) = P(sin^2\frac{\omega}{2})$ in equation (3.39) by retaining the zeros within the unit circle [98] [99]. For each L, the corresponding ${}_LH(\omega)$ has $2L$ zeros which are concentrated at $\omega = \pi$ and the resulting $H(\omega)$ will have exactly $N = 2L$ coefficients. This leads to the choice of phase of ${}_LH(\omega)$ so that [98],

$${}_LH(\omega) = \frac{1}{\sqrt{2}} \sum_{n=0}^{2L-1} h(n) e^{-in\omega}. \tag{3.42}$$

In order to make smooth wavelet bases, the number of vanishing moments L of $\psi(x)$ can be increased. From equation (3.33), the multiplicity of zeros at $\omega = 0$ in $G(\omega)$ is also L, and hence the length of $g(n)$. That is,

$$\int x^l \psi(x) dx = 0 \text{ for } l = 0, 1, ...L-1 \tag{3.43}$$

or equivalently

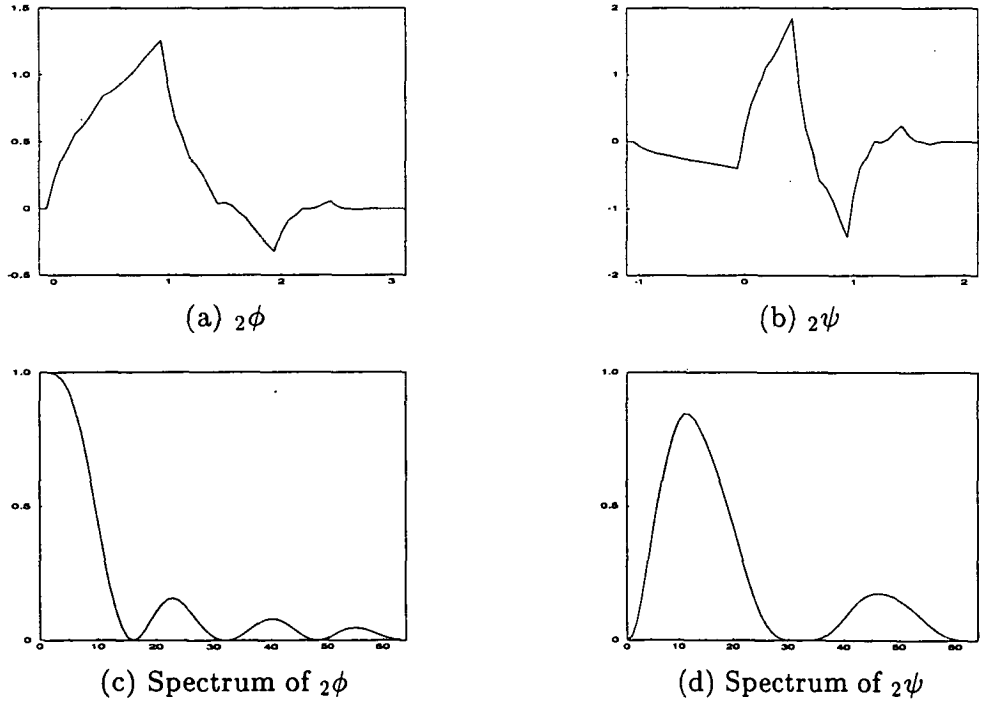
$$\frac{d^l}{d\omega^l} \hat{\psi}(\omega)|_{\omega=0} = 0 \quad \text{for} \quad l \leq L-1 \tag{3.44}$$

where $\psi \in C^{L-1}$. This means that the regularity is increased with an increasing number of vanishing moments L in $\psi(x)$.

The filter coefficients $h(n)$ for $L = 2, 3, 4$ and $L = 6$ are tabulated in Table 3.1. Figures 3.6, 3.7, 3.8 and 3.9 show the scaling functions ϕ and mother wavelets ψ, and their spectra for $L = 2, 3, 4$ and $L = 6$ respectively.

L	n	$h(n)$	L	n	$h(n)$
2	0	0.4829629	3	0	0.3326705
	1	0.8365163		1	0.8068915
	2	0.2241439		2	0.4598775
	3	-0.1294095		3	-0.1350110
				4	-0.0854413
				5	0.0352263
4	0	0.2303778	6	0	0.1115407
	1	0.7148466		1	0.4946239
	2	0.6308808		2	0.7511339
	3	-0.0279838		3	0.3152503
	4	-0.1870348		4	-0.2262647
	5	0.0308414		5	-0.1297669
	6	0.0328830		6	0.0975016
	7	-0.0105974		7	0.0275229
				8	-0.0315820
				9	0.0005538
				10	0.0047772
				11	-0.0010773

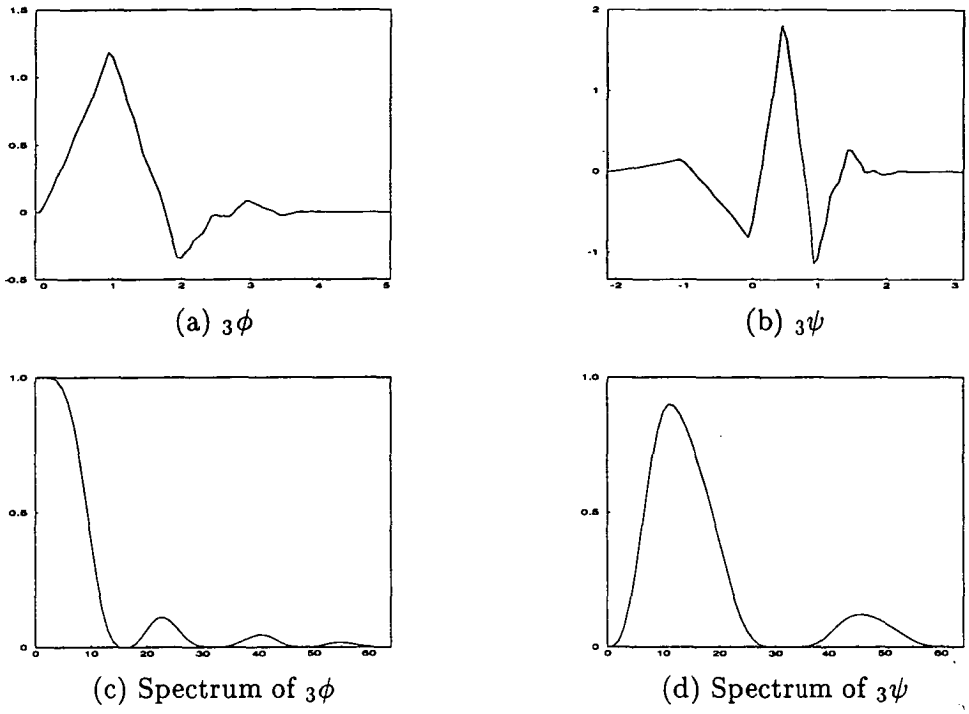
Table 3.1: Filter coefficients $h(n)$ for compactly supported Daubechies wavelet bases for $L = 2, 3, 4$ and 6.



(a) ${}_2\phi$ (b) ${}_2\psi$

(c) Spectrum of ${}_2\phi$ (d) Spectrum of ${}_2\psi$

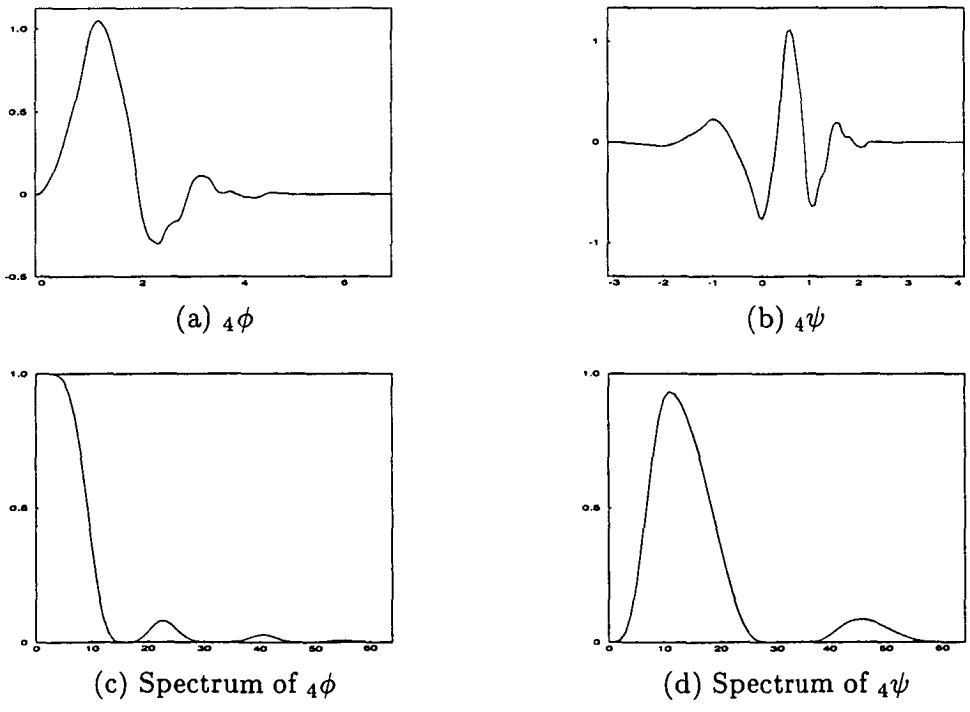
Figure 3.6: The scaling functions ${}_2\phi$ and wavelets ${}_2\psi$ and their spectra for the compactly supported wavelets with a maximum number of vanishing moments.



(a) $_3\phi$ (b) $_3\psi$

(c) Spectrum of $_3\phi$ (d) Spectrum of $_3\psi$

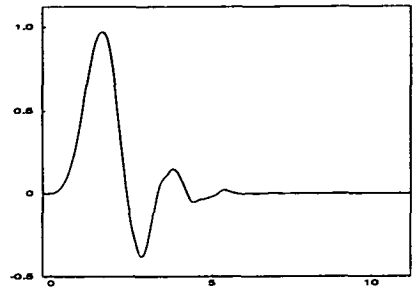
Figure 3.7: The scaling functions $_2\phi$ and wavelets $_2\psi$ and their spectra for the compactly supported wavelets with a maximum number of vanishing moments.



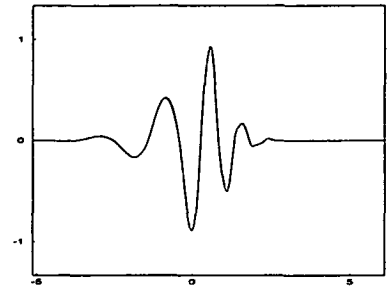
(a) $_4\phi$ (b) $_4\psi$

(c) Spectrum of $_4\phi$ (d) Spectrum of $_4\psi$

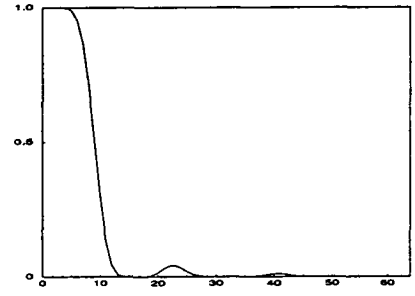
Figure 3.8: The scaling functions $_4\phi$ and wavelets $_4\psi$ and their spectra for the compactly supported wavelets with a maximum number of vanishing moments.



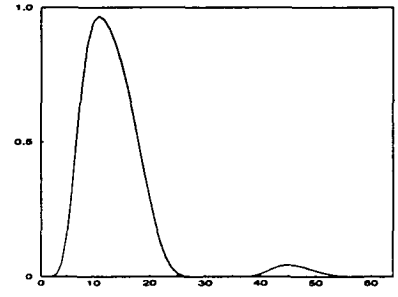
(a) ${}_6\phi$



(b) ${}_6\psi$



(c) Spectrum of ${}_6\phi$



(d) Spectrum of ${}_6\psi$

Figure 3.9: The scaling functions ${}_6\phi$ and wavelets ${}_6\psi$ and their spectra for the compactly supported wavelets with a maximum number of vanishing moments.

It can be seen that the family of Daubechies wavelets are an orthonormal set that are compactly supported, but asymmetric. They are supported on an interval of length $2L - 1$.

Maximum Vanishing Moment Wavelets

For $\mathcal{R} = 0$ in equation (3.41), the resulting $H(\omega)$ has exactly $N = 2L$ coefficients and the $2L$ zeros in equation (3.39) are concentrated at $\omega = \pi$. Therefore, $H(\omega)$ will have a maximum number of L vanishing moments. If we systematically retain all L zeros of $P(y)$ within the unit circle in the factorization, we then have a minimum phase filter. This does correspond however, to a very markedly asymmetrical ϕ and ψ.

For $\mathcal{R} \neq 0$, we can move some of the zeros away from $\omega = \pi$ in equation (3.39) and put them into $P(y)$. The number of vanishing moments consequently decreases, but now we have more choices in the factorization of $P(y)$. If we only choose those complex and real zeros so that $H(\omega)$ has as close to linear phase as possible, the resulting ϕ and ψ will be least asymmetrical but less smooth.

The maximum vanishing moment can also be derived from the subband filtering scheme as described in the previous section. From equation (3.27), we obtain $N/2$ equations. We require $N/2$ more equations to solve for the N coefficients $h(0), h(1), ..., h(N-1)$. If we aim to design $\psi(x)$, and hence $g(n)$, to have a maximum number of vanishing moments, then $L = N/2$. We therefore have the $N/2$ equations from

$$\sum_n n^l g(n) = 0 \text{ for } l = 0, 1, ..., L-1. \tag{3.45}$$

The family of Daubechies's filters therefore has a maximum number of vanishing moments.

Regularity of Compactly Supported Orthonormal Wavelets

An orthonormal basis of wavelets is in C^{L-1}, i.e. $\psi_{j,k} \leq C(1+|x|)^{-L}$, only if the mother wavelet has L vanishing moments. For compactly supported orthonormal wavelets regularity is not easy to determine because of the behaviour of $\mathcal{L}(\omega)$ in equation (3.39) [99]. In general, they have a non-integer Hölder exponent and are more regular at some points than others in the wavelets. It has been shown [99] that the exact asymptotic decay of ${}_L\hat{\phi}(\omega)$ is

$$\left|{}_L\hat{\phi}(\omega)\right| \leq C\,(1+|\omega|)^{-L+\log_2 B} \tag{3.46}$$

where

$$B = \sup|Z_L(\omega)| \leq 2^{L-1} \qquad \omega \in [0, \pi] \tag{3.47}$$

for $\hat{\phi}(\omega)$ to decay sufficiently to produce smoothness in $\phi(x)$. For a particular choice of ϕ and ψ corresponding to each L, Daubechies [99] was able to establish that the regularity of ${}_L\phi$ and ${}_L\psi$ increased linearly with L and that ${}_L\phi, {}_L\psi \in C^{\alpha L}$ where $\alpha \approx 0.2075$. This means that almost 80 % of the zero moments are not contributing to the regularity of the compactly supported orthonormal wavelets. The regularity estimates for smaller values of L are tabulated in Table 3.2 which guarantee the asymptotic decay of ${}_L\phi$ [99].

3.6.2 Near Linear Phase Orthonormal Wavelet Bases

The construction of ${}_L\phi, {}_L\psi$ as described above is derived from equation (3.39), and can be rewritten as,

$$|{}_LH(\omega)|^2 = \left(cos\frac{\omega}{2}\right)^{2L} P_L(sin^2\frac{\omega}{2}),$$

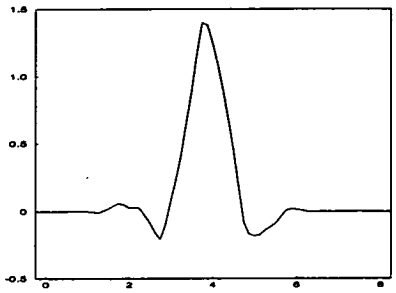
L	α
2	0.500
3	0.915
4	1.275
5	1.596
6	1.888
8	2.415

Table 3.2: The regularity estimation of ${}_L\phi \in C^\alpha$.

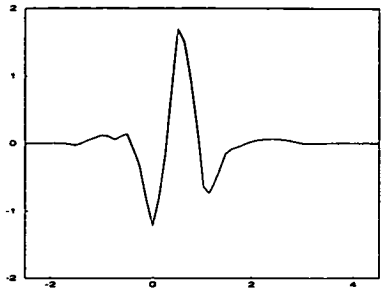
where the coefficients ${}_Lh(n)$ are determined by taking the square root of P_L via spectral factorization. Factorization of all L zeros of $P_L(sin^2\frac{\omega}{2})$ within the unit circle results in a minimum phase filter. If we only choose those complex and real zeros so that $H(\omega)$ has as close to linear phase as possible, the resulting ϕ and ψ will be least asymmetrical but less smooth. Daubechies [99] has pointed out that a close to linear phase $H(\omega)$ can be obtained by retaining the zeros from every quaduplet or duplet in which the total phase is as close to zero as possible. In practice, we have $2^{L/2}$ choices and this number can be reduced by another factor of 2 for every choice [99]. For $L < 4$, there is only one pair ϕ_L, ψ_L. For $L \geq 4$ there are more pairs that are close to linear phase. Figures 3.10 and 3.11 show the construction of ϕ and ψ and their spectra, which have close to linear phase for $L = 4$ and $L = 6$ respectively. Their filter coefficients are tabulated in Table 3.3.

L	n	$h(n)$	L	n	$h(n)$
4	0	-0.075766	6	0	0.015404
	1	-0.029636		1	0.003491
	2	0.497619		2	-0.117990
	3	0.803739		3	-0.048312
	4	0.297858		4	0.491056
	5	-0.099220		5	0.787641
	6	-0.012604		6	0.337929
	7	0.032223		7	-0.072638
				8	-0.021060
				9	0.044725
				10	0.001768
				11	-0.007801

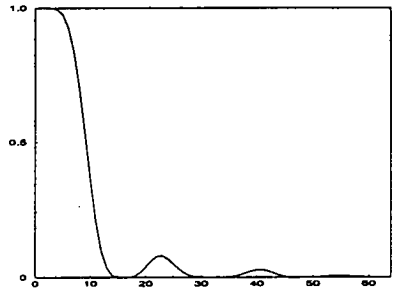
Table 3.3: Filter coefficients $h(n)$ for near linear phase wavelet bases for $L = 4$ and $L = 6$.



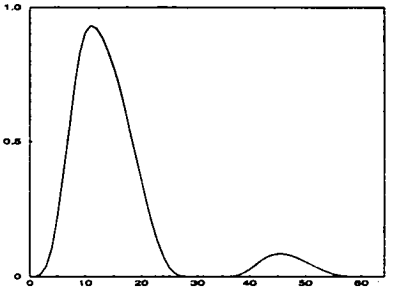
(a) ${}_4\phi$



(b) ${}_4\psi$

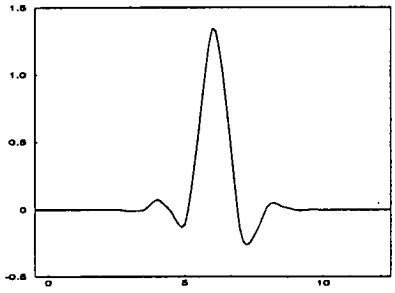


(c) Spectrum of ${}_4\phi$

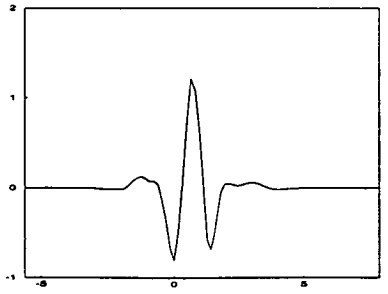


(d) Spectrum of ${}_4\psi$

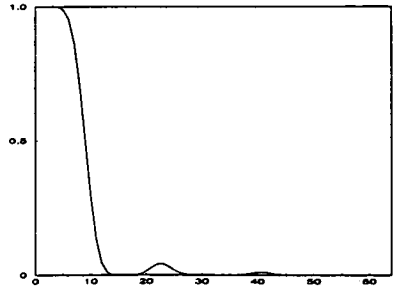
Figure 3.10: The scaling functions ${}_4\phi$ and wavelets ${}_4\psi$ and their spectra for near linear phase wavelet bases.



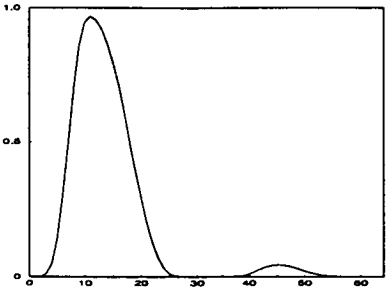
(a) ${}_6\phi$



(b) ${}_6\psi$



(c) Spectrum of ${}_6\phi$



(d) Spectrum of ${}_6\psi$

Figure 3.11: The scaling functions ${}_6\phi$ and wavelets ${}_6\psi$ and their spectra for near linear phase wavelet bases.

3.6.3 Least Asymmetrical Orthonormal Wavelet Bases

The conventional requirement for a smooth orthonormal basis is that the wavelet ψ has to have as many vanishing moments as possible. An exact requirement on ϕ cannot be satisfied since $\int \phi(x)dx = 1$. Coifman [109] constructed orthonormal wavelet bases with vanishing moments not only for ψ but also for ϕ. These wavelets are known as "coiflets" and have maximum regularity in both ψ and ϕ. The coiflet needs to satisfy two conditions, i.e.,

$$\int x^l \psi(x)dx = 0 \qquad l = 0, ..., L-1 \tag{3.48}$$

and

$$\int \phi(x) = 1, \quad \int x^l \phi(x)dx = 0 \quad l = 1, ..., L-1 \tag{3.49}$$

where L is the order of coiflet. Equation (3.48) can be expressed in terms of $H(\omega)$, where

$$H(\omega) = \left(\frac{1+e^{-i\omega}}{2}\right)^L \tilde{Z}_L(\omega). \tag{3.50}$$

From equation (3.49), we obtain the condition

$$\frac{d^l}{d\omega^l}\hat{\phi}(\omega)|_{\omega=0} = 0 \qquad l = 1, ..., L-1. \tag{3.51}$$

Recalling equation (3.35) and substituting into equation (3.51), we then have

$$\hat{\phi}'(\omega) = \frac{1}{2}H'(\frac{\omega}{2})\hat{\phi}(\frac{\omega}{2}) + \frac{1}{2}H(\frac{\omega}{2})\hat{\phi}'(\frac{\omega}{2}) \tag{3.52}$$

and hence

$$H'(0) = \sqrt{(2\pi)}\hat{\phi}'(0). \tag{3.53}$$

Consequently, $\int x\phi(x)dx = 0$ is equivalent to $H'(0) = 0$. From equation (3.49), we then have

$$H(\omega) = 1 + (1 - e^{-i\omega})^L \tilde{Z}_L(\omega) \tag{3.54}$$

where $\tilde{Z}_L$ is a trigonometric polynomial.

Equations (3.50) and (3.54) have to satisfy equation (3.31). This leads to the choice of L being even [99]. Equations (3.50) and (3.54) imply that we have two trigonometric

polynomials $\mathcal{P}_1$ and $\mathcal{P}_2$, such that,

$$\left(cos^2\frac{\omega}{2}\right)^K \mathcal{P}_1(\omega) = 1 + \left(sin^2\frac{\omega}{2}\right)^K \mathcal{P}_2(\omega) \tag{3.55}$$

where $2K = L$. This equation has the form of Bezout's equation [99],

$$(1-y)^K \mathcal{P}_1(y) = 1 + y^K \mathcal{P}_2(1-y) \tag{3.56}$$

where $y = sin^2\frac{\omega}{2}$, and the solution is

$$\mathcal{P}_1(\omega) = \sum_{k=0}^{K-1} \begin{pmatrix} K-1+k \\ k \end{pmatrix} y^k + y^K r(\omega) \tag{3.57}$$

where r is an arbitrary trigonometric polynomial, and is tailored so as to satisfy the orthonormality requirement in equation (3.31). The complex function of $r(\omega)$ is defined by [99]

$$r(\omega) = \sum_{n=0}^{2K-1} r(n) e^{-in\omega}. \tag{3.58}$$

Therefore, $H(\omega)$ can be defined by

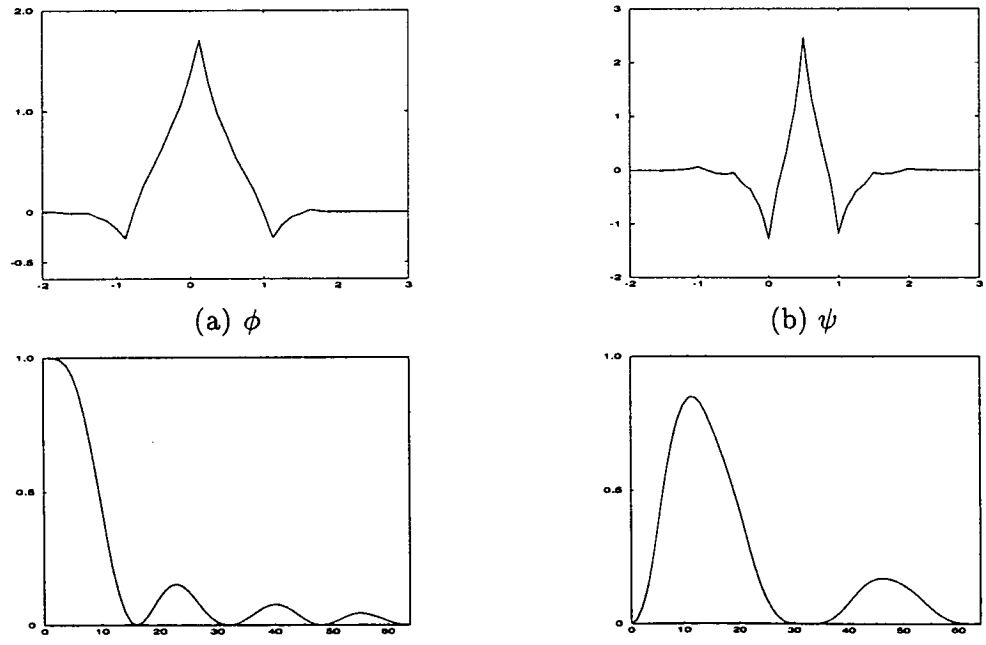
$$H(\omega) = \left(\frac{1+e^{-in\omega}}{2}\right)^{2K} \mathcal{P}_1(\omega). \tag{3.59}$$

Small values of K result in the smoothest coiflets, but they are not the most symmetric ones. For $K = 1$, there exists a less asymmetric coiflet with compact support. The filter coefficients of $h(n)$ are tabulated in Table 3.4 for $L = 2$ and $L = 4$. Figures 3.12 and 3.13 show the coiflets and their spectra for $L = 2$ and $L = 4$ respectively.

The obvious advantage of coiflets is that their ϕ and ψ are both smooth and much more symmetric and similar to one another than other compactly supported orthonormal wavelet bases. The price to pay for this advantage is that a coiflet with L vanishing moments typically has a support width $(3L-1)$ compared to the conventional orthonormal wavelet bases which have a support width $(2L-1)$. Coiflets therefore do not have a maximum number of vanishing moments.

L	n	$h(n)$	L	n	$h(n)$
2	0	-0.072733	4	0	0.016387
	1	0.337898		1	-0.041465
	2	0.852572		2	-0.067373
	3	0.384865		3	0.386110
	4	-0.072733		4	0.812724
	5	-0.015656		5	0.417005
				6	-0.076489
				7	-0.059434
				8	0.023680
				9	0.005611
				10	-0.001823
				11	-0.000721

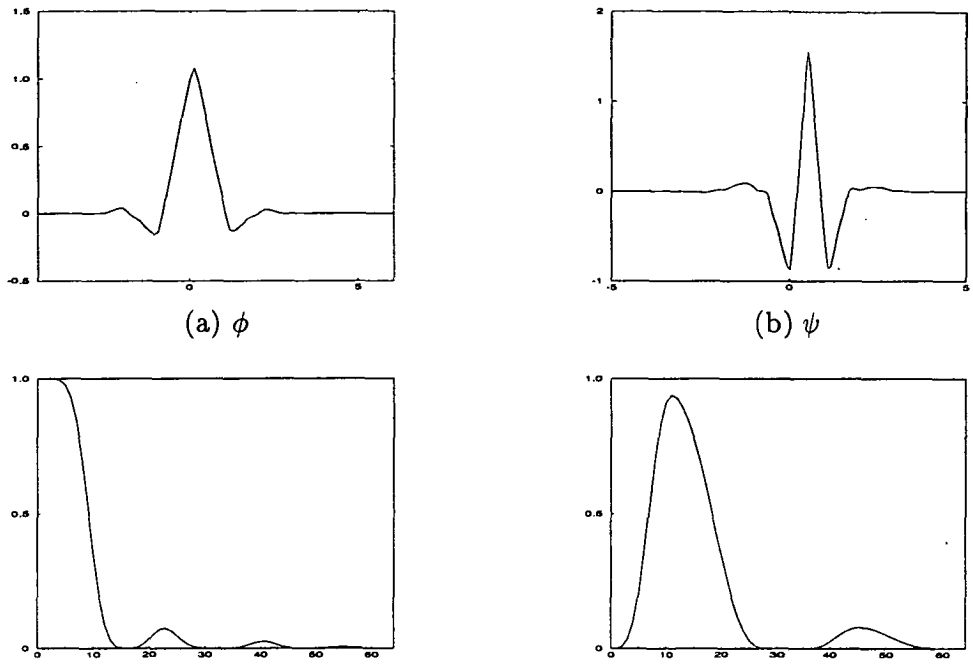
Table 3.4: Filter coefficients $h(n)$ for coiflets with $L = 2$ and $L = 4$.



(a) ϕ (b) ψ

(c) Spectrum of ϕ (d) Spectrum of ψ

Figure 3.12: The coiflet and scaling functions for $L = 2$ and their spectra.



(a) ϕ (b) ψ

(c) Spectrum of ϕ (d) Spectrum of ψ

Figure 3.13: The coiflet and scaling functions for $L = 4$ and their spectra.

3.6.4 Design of Compactly Supported Biorthonormal Wavelet Bases

Biorthonormal wavelet bases of finite support can be designed for exact reconstruction if we no longer require the lengths of the analysis and synthesis filters to be the same. If $H(\omega)$ is chosen to be a linear phase filter and symmetric, i.e. $H(-\omega) = e^{2ik\omega}H(\omega)$, then $H(\omega)$ can be rewritten as,

$$H(\omega) = e^{-ik\omega}p_0(cos\omega) \tag{3.60}$$

where p_0 is a polynomial. Let $\tilde{H}(\omega)$ be chosen to have the same form, then

$$\tilde{H}(\omega) = e^{-ik\omega}\tilde{p}_0(cos\omega) \tag{3.61}$$

where $\tilde{p}_0$ is any polynomial that satisfies

$$p_0(y)\overline{\tilde{p}_0(y)} + p_0(-y)\overline{\tilde{p}_0(-y)} = 1. \tag{3.62}$$

For biorthogonality [100] the filters G and $\tilde{G}$ are related to H and $\tilde{H}$ by

$$\begin{aligned} G(\omega) &= e^{-i\omega}\overline{\tilde{H}(\omega+\pi)} \\ \text{and} \quad \tilde{G}(\omega) &= e^{-i\omega}\overline{H(\omega+\pi)}. \end{aligned} \tag{3.63}$$

Recalling the perfect reconstruction property, it can be shown that

$$\begin{aligned} H(\omega)\overline{\tilde{H}(\omega)} + H(\omega+\pi)\overline{\tilde{H}(\omega+\pi)} &= 1 \\ \text{and} \quad H(\omega)\overline{\tilde{H}(\omega)} + G(\omega)\overline{\tilde{G}(\omega)} &= 1. \end{aligned} \tag{3.64}$$

Since $H(\omega)$ and $\tilde{H}(\omega)$ have a multiple zero at $\omega = \pi$, $p_0(cos\omega)$ and $\tilde{p}_0(cos\omega)$ are divisible by $(\frac{1+e^{-i\omega}}{2})^L$ for $L > 0$. But in order to obtain a polynomial of $cos(\omega)$, L has to be even. Therefore $H(\omega)$ and $\tilde{H}(\omega)$ have the general form denoted by

$$\begin{aligned} H(\omega) &= e^{-ik\omega/2}\,(cos\tfrac{\omega}{2})^{2l}\,q_0(cos\omega) \\ \tilde{H}(\omega) &= e^{-ik\omega/2}\,(cos\tfrac{\omega}{2})^{2\tilde{l}}\,\tilde{q}_0(cos\omega) \end{aligned} \tag{3.65}$$

for an even number of taps, where $2l = L$ and $2\tilde{l} = L$, or

$$\begin{aligned} H(\omega) &= e^{-ik\omega/2}\,(cos\tfrac{\omega}{2})^{2l+1}\,q_0(cos\omega) \\ \text{and} \quad \tilde{H}(\omega) &= e^{-ik\omega/2}\,(cos\tfrac{\omega}{2})^{2\tilde{l}+1}\,\tilde{q}_0(cos\omega) \end{aligned} \tag{3.66}$$

for an odd number of taps, where $2l + 1 = L$ and $2\tilde{l} + 1 = \tilde{L}$; and $k = 0$ if l and $\tilde{l}$ are even and $k = 1$ if l and $\tilde{l}$ are odd. Substitution of equations (3.65) and (3.66) into equation (3.64) gives

$$\left(cos\frac{\omega}{2}\right)^{2L} q_0(cos\omega)\overline{\tilde{q}_0(cos\omega)} + \left(sin\frac{\omega}{2}\right)^{2L} q_0(-cos\omega)\overline{\tilde{q}_0(-cos\omega)} = 1 \tag{3.67}$$

where $L = l + \tilde{l}$ for an even number of taps and $L = l + \tilde{l} + 1$ for an odd number of taps. Equation (3.67) has the form of Bezout's equation and the solution is denoted by

$$P(y) = \sum_{k=0}^{L-1} \begin{pmatrix} L-1+k \\ k \end{pmatrix} y^k + y^L \mathcal{R}(\frac{1}{2} - y) \tag{3.68}$$

for $P(y) = q_0(cos\omega)\overline{\tilde{q}_0(cos\omega)}$.

B-Spline Bases

We can choose $q_0(cos\omega)$ or $\tilde{q}_0(cos\omega)$ and $\mathcal{R}(cos\omega)$ in equation (3.68) to simplify the design and to obtain other characteristics apart from being symmetric. If we choose $\mathcal{R} \equiv 0$ and $\tilde{q}_0 \equiv 1$, we then have

$$\begin{aligned} \tilde{H}(\omega) &= (cos\tfrac{\omega}{2})^L \\ \text{or} \quad \tilde{H}(\omega) &= (cos\tfrac{\omega}{2})^{L+1} \end{aligned} \tag{3.69}$$

where $L = 2l$ or $L = 2l + 1$. Therefore,

$$H(\omega) = \left(cos\frac{\omega}{2}\right)^L \sum_{k=0}^{l+\tilde{l}-1} \begin{pmatrix} l+\tilde{l}-1+k \\ k \end{pmatrix} \left(sin^2\frac{\omega}{2}\right)^k \tag{3.70}$$

for $\tilde{L} = 2\tilde{l}$ or

$$H(\omega) = e^{-i\omega/2} \left(cos\frac{\omega}{2}\right)^L \sum_{k=0}^{l+\tilde{l}} \begin{pmatrix} l+\tilde{l}+k \\ k \end{pmatrix} \left(sin^2\frac{\omega}{2}\right)^k \tag{3.71}$$

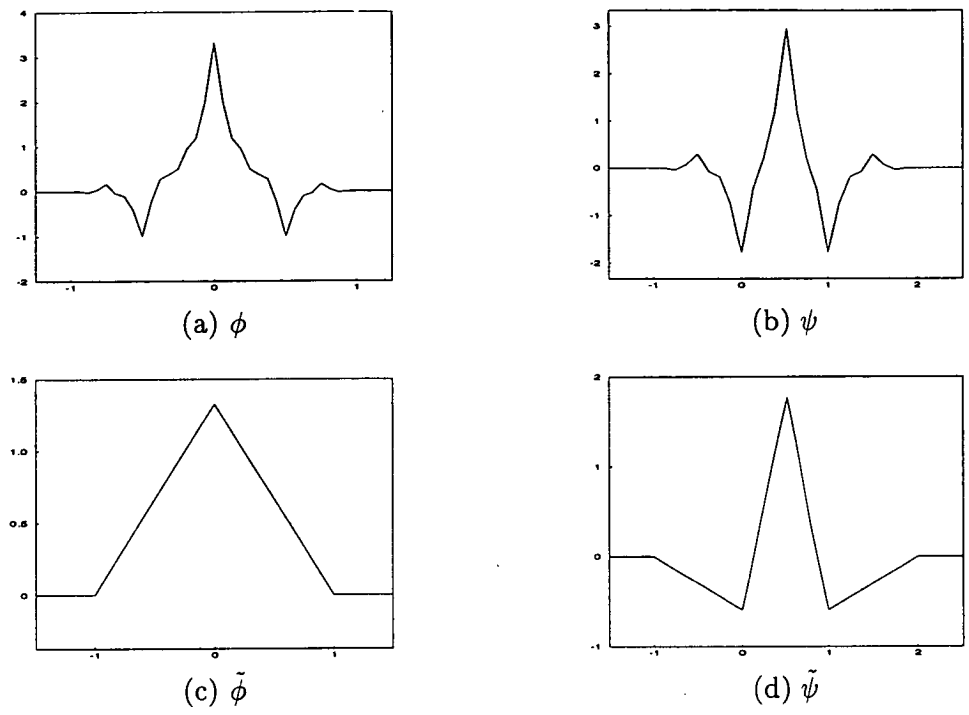
for $\tilde{L} = 2\tilde{l} + 1$ respectively. We can choose l freely. The construction of wavelet bases with these properties results in a family of biorthonormal bases in which ψ is a spline function of compact support. This wavelet basis is also known as the "Battle-Lemarie" family of B-spline wavelets. The shape of $\tilde{\psi}$ is determined mainly by $\tilde{L}$ and is changed only slightly by L. ϕ and ψ on the other hand depend on both L and $\tilde{L}$. For $\tilde{L} = 2$, $L = 2$ and $L = 4$, the filter pairs for decomposition and reconstruction are tabulated in Table 3.5 and Table 3.6. Figures 3.14, 3.15, 3.16 and 3.17 show the scaling and wavelet functions for the biorthonormal wavelet bases and their spectra for the construction of spline bases, for $\tilde{L} = 2$, $L = 2$ and $\tilde{L} = 2$, $L = 4$ respectively.

n	$h(n)$	$g(n)$	$\tilde{h}(n)$	$\tilde{g}(n)$
-3	0	0	0	$-\frac{1}{8}\sqrt{2}$
-2	$-\frac{1}{8}\sqrt{2}$	$-\frac{1}{4}\sqrt{2}$	0	$-\frac{1}{4}\sqrt{2}$
-1	$\frac{1}{4}\sqrt{2}$	$\frac{1}{2}\sqrt{2}$	$\frac{1}{4}\sqrt{2}$	$\frac{3}{4}\sqrt{2}$
0	$\frac{3}{4}\sqrt{2}$	$-\frac{1}{4}\sqrt{2}$	$\frac{1}{2}\sqrt{2}$	$-\frac{1}{4}\sqrt{2}$
1	$\frac{1}{4}\sqrt{2}$	0	$\frac{1}{4}\sqrt{2}$	$-\frac{1}{8}\sqrt{2}$
2	$-\frac{1}{8}\sqrt{2}$	0	0	0
3	0	0	0	0

Table 3.5: The B-spline filter coefficients $h(n)$, $g(n)$, $\tilde{h}(n)$ and $\tilde{g}(n)$ for $\tilde{L} = 2$ and $L = 2$.

n	$h(n)$	$g(n)$	$\tilde{h}(n)$	$\tilde{g}(n)$
-5	0	0	0	$\frac{3}{128}\sqrt{2}$
-4	$\frac{3}{128}\sqrt{2}$	0	0	$\frac{3}{64}\sqrt{2}$
-3	$-\frac{3}{64}\sqrt{2}$	0	0	$\frac{1}{8}\sqrt{2}$
-2	$-\frac{1}{8}\sqrt{2}$	$-\frac{1}{4}\sqrt{2}$	0	$-\frac{19}{64}\sqrt{2}$
-1	$\frac{19}{64}\sqrt{2}$	$\frac{1}{2}\sqrt{2}$	$\frac{1}{4}\sqrt{2}$	$\frac{45}{64}\sqrt{2}$
0	$\frac{45}{64}\sqrt{2}$	$-\frac{1}{4}\sqrt{2}$	$\frac{1}{2}\sqrt{2}$	$-\frac{19}{64}\sqrt{2}$
1	$\frac{19}{64}\sqrt{2}$	0	$\frac{1}{2}\sqrt{2}$	$\frac{1}{8}\sqrt{2}$
2	$-\frac{1}{8}\sqrt{2}$	0	0	$\frac{3}{64}\sqrt{2}$
3	$-\frac{3}{64}\sqrt{2}$	0	0	$\frac{3}{128}\sqrt{2}$
4	$\frac{3}{128}\sqrt{2}$	0	0	0
5	0	0	0	0

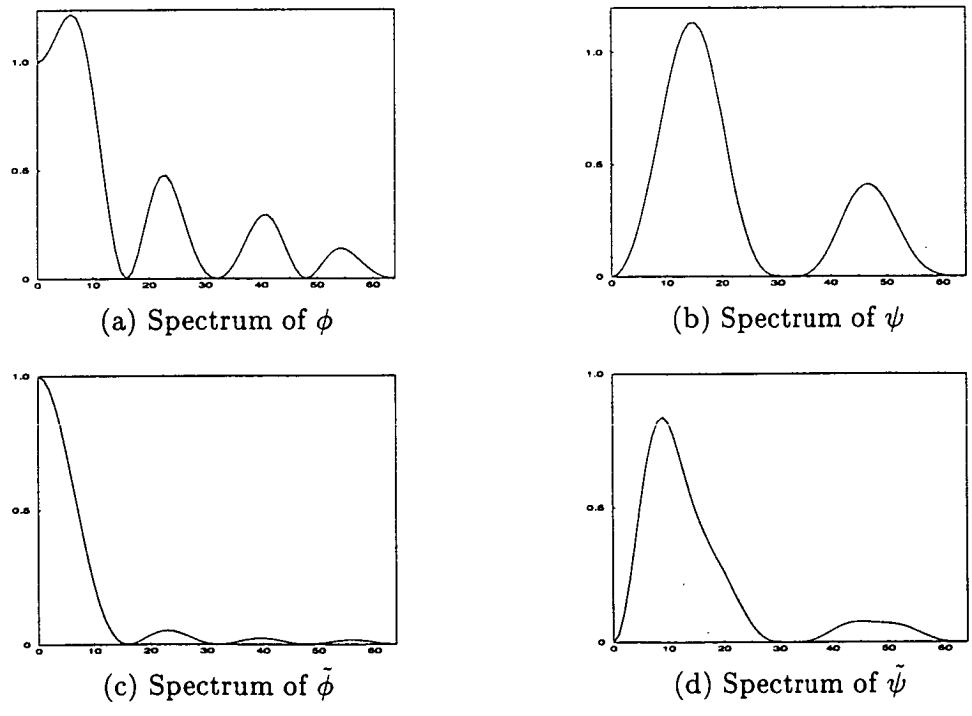
Table 3.6: The B-spline filter coefficients $h(n)$, $g(n)$, $\tilde{h}(n)$ and $\tilde{g}(n)$ for $L = 4$ and $\tilde{L} = 2$.



(a) ϕ (b) ψ

(c) $\tilde{\phi}$ (d) $\tilde{\psi}$

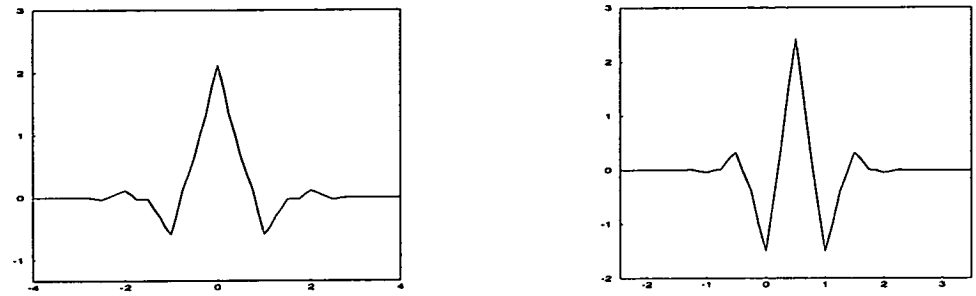
Figure 3.14: The B-spline scaling and wavelet functions for $\tilde{L} = 2$ and $L = 2$. (a) and (b) for analysis, (c) and (d) for synthesis.



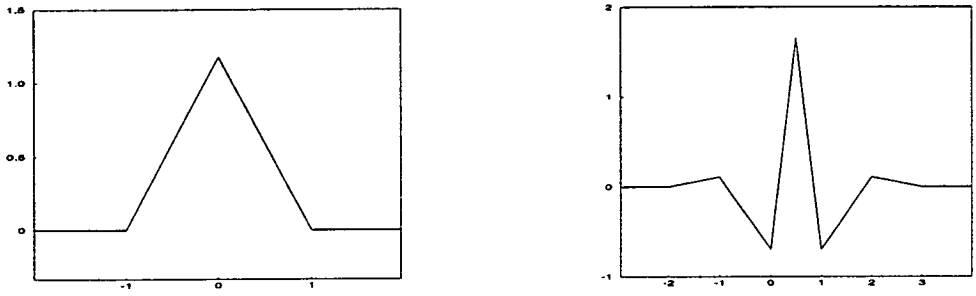
(a) Spectrum of ϕ (b) Spectrum of ψ

(c) Spectrum of $\tilde{\phi}$ (d) Spectrum of $\tilde{\psi}$

Figure 3.15: The spectra of the B-spline scaling and wavelet functions for $\tilde{L} = 2$ and $L = 2$. (a) and (b) for analysis, (c) and (d) for synthesis.

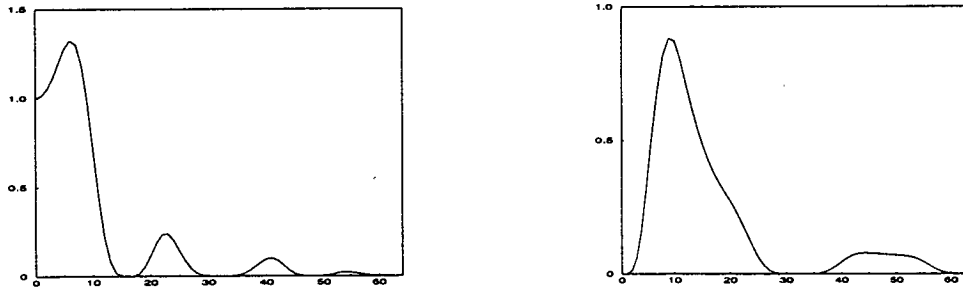


(a) ϕ (b) ψ

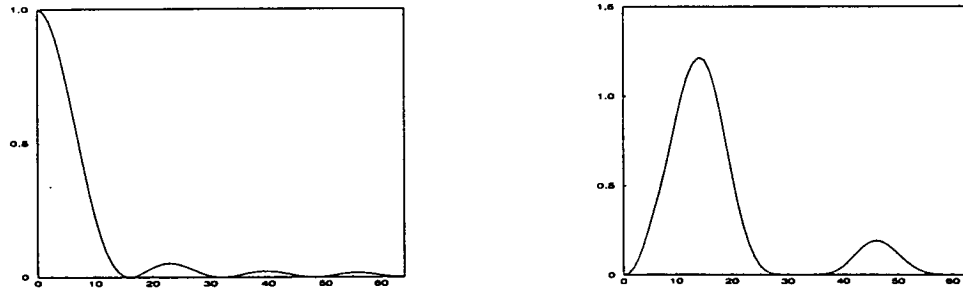


(c) $\tilde{\phi}$ (d) $\tilde{\psi}$

Figure 3.16: The B-spline scaling and wavelet functions for $\tilde{L} = 2$ and $L = 4$. (a) and (b) for analysis, (c) and (d) for synthesis.



(a) Spectrum of ϕ (b) Spectrum of ψ



(c) Spectrum of $\tilde{\phi}$ (d) Spectrum of $\tilde{\psi}$

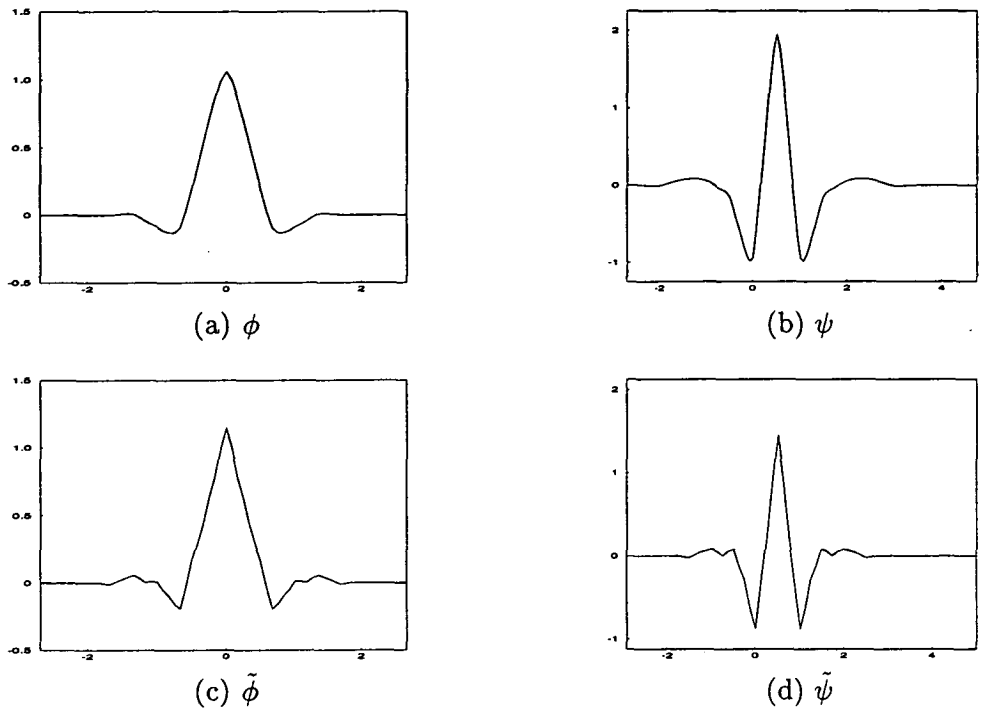
Figure 3.17: The spectra of the B-spline scaling and wavelet functions for $\tilde{L} = 2$ and $L = 4$. (a) and (b) for analysis, (c) and (d) for synthesis.

The greatest advantage of the B-spline wavelet is that their filter coefficients are all rational being fractional powers of 2 and are therefore most efficient for fast computation and hardware implementations. Furthermore, spline wavelets ψ and scaling functions ϕ are known exactly in closed forms. The main disadvantage of spline wavelet bases is in their very unequal support widths for ϕ and $\tilde{\phi}$. The support widths of ψ and $\tilde{\psi}$ are always equal and is given by $(L + \tilde{L} - 1)$.

In order to produce a less disparate filter length for the spline wavelet bases, the polynomial of degree $l + \tilde{l} - 1$ in $sin(\frac{\omega}{2})$ can be broken up into a product of polynomials $sin(\frac{\omega}{2})$ with real coefficients [99] [108]. One to be allocated to H and the other to $\tilde{H}$. There are other ways of making the length difference of H and $\tilde{H}$ as small as possible, as demonstrated by Daubechies in [99]. There is, however, a limit on how small the filters can be since the minimum value of $l + \tilde{l}$ is equal to 4 [99]. The filter coefficients for filters with similar lengths, however, are no longer rational numbers. The B-spline filter coefficients for $L + \tilde{L} = 8$ where the filters have similar lengths are tabulated in Table 3.7. Figures 3.18 and 3.19 show the scaling function and mother wavelet and their spectra for analysis and synthesis respectively. It can be seen that the shape of ϕ and $\tilde{\phi}$, and ψ and $\tilde{\psi}$ are very similar.

n	$h(n)$	$g(n)$	$\tilde{h}(n)$	$\tilde{g}(n)$
-5	0	0	0	0
-4	0.000000	0.023849	0.037829	0.064539
-3	-0.064539	-0.110624	-0.023849	-0.040690
-2	-0.040690	-0.377403	-0.110624	-0.418092
-1	0.418092	0.852699	0.377403	0.788485
0	0.788485	-0.377403	0.852699	-0.418092
1	0.418092	-0.110624	0.377403	-0.040690
2	-0.040690	0.023849	-0.110624	0.064539
3	-0.064539	0.037829	-0.023849	0
4	0	0	0.037829	0
5	0	0	0	0

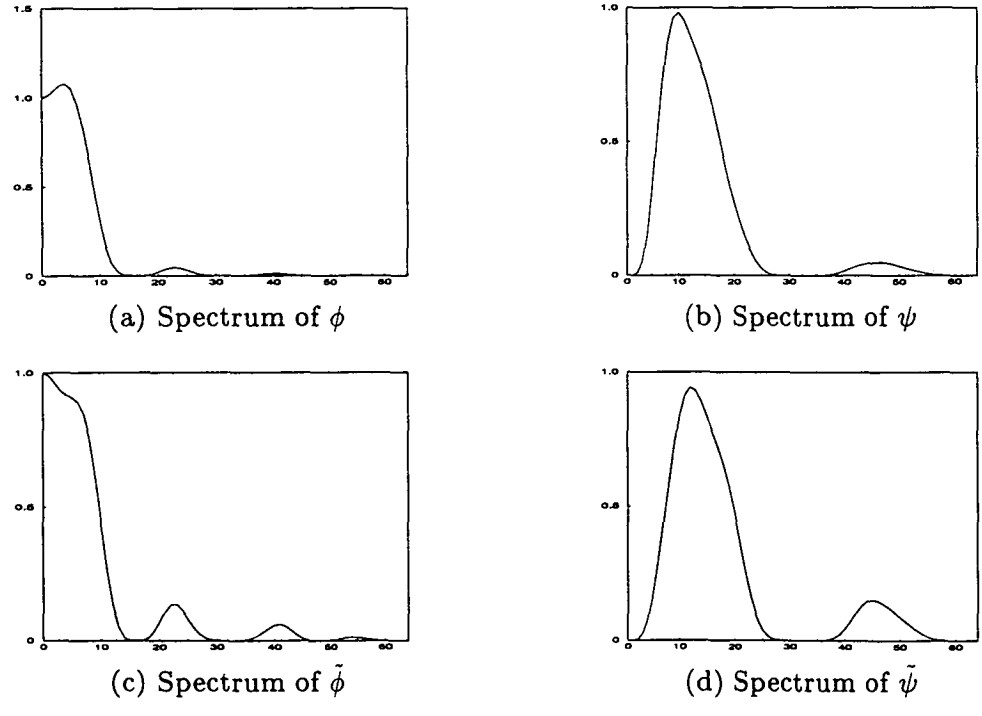
Table 3.7: The B-spline filter coefficients $h(n)$, $g(n)$, $\tilde{h}(n)$ and $\tilde{g}(n)$ for $\tilde{L} + L = 8$ with filters of similar length.



(a) ϕ (b) ψ

(c) $\tilde{\phi}$ (d) $\tilde{\psi}$

Figure 3.18: The B-spline scaling and wavelet functions with filters of similar length for $\tilde{L} + L = 8$. (a) and (b) for analysis, (c) and (d) for synthesis.



(a) Spectrum of ϕ (b) Spectrum of ψ

(c) Spectrum of $\tilde{\phi}$ (d) Spectrum of $\tilde{\psi}$

Figure 3.19: The spectra of the B-spline scaling and wavelet functions with filters of similar length for $\tilde{L} + L = 8$. (a) and (b) for analysis, (c) and (d) for synthesis.

Biorthonormal Wavelet Bases close to an Orthonormal Basis

Among the choices of biorthonormal wavelets for which $\mathcal{R} \neq 0$ in equation (3.68) there exists a special choice of $\mathcal{R}$ for which the analysis and synthesis filters are very close to each other and both very close to an orthonormal filter. The popular Laplacian pyramid filter [45] can serve as either the analysis or synthesis filter for the biorthonormal wavelet basis. These filters are given explicitly by

$$-ae^{-2i\omega} + 0.25e^{-i\omega} + (0.50 + 2a) + 0.25e^{i\omega} - ae^{2i\omega}.$$

We can then choose the Laplacian pyramid filter as the analysis filter of a biorthonormal wavelet basis i.e.,

$$H(\omega) = \left(cos\frac{\omega}{2}\right)^2 \left[1 + 16asin^2\frac{\omega}{2}\right]. \tag{3.72}$$

The corresponding synthesis filter, $\tilde{H}(\omega)$, which satisfies the biorthonormality and exact reconstruction requirement in equation (3.32), should also satisfy Bezout's equation. Therefore, $\tilde{H}(\omega)$ can be rewritten as

$$\tilde{H}(\omega) = (\cos\frac{\omega}{2})^2 \left[1 + 2(1-8a)sin^2\frac{\omega}{2} - 32a\frac{1-8a}{1+8a}sin^4\frac{\omega}{2}\right]. \tag{3.73}$$

It can be readily verified that for $a = 0$ and $a = -\frac{1}{16}$, $H(\omega)$ reduces to a spline filter of order 2 and 4 respectively [1] [2]. If we apply the required regularity condition on B in equation (3.47) for $Z_l(\omega)$ of $H(\omega)$ and $\tilde{H}(\omega)$ in equation (3.72) and equation (3.73), we see that equation (3.72) satisfies this condition when $-0.1875 \leq a \leq 0.0625$, and equation (3.73) satisfies the same requirement when $0.0275 \leq a \leq 0.285$. Consequently, usable values of a to obtain smoothness in both $\phi(x)$ and $\tilde{\phi}(x)$ are found in the small range $0.0275 \leq a \leq 0.0625$. Since $\phi(x)$ becomes smoother as a becomes progressively smaller than 0.0625, while $\tilde{\phi}(x)$ becomes smoother as a becomes progressively greater than 0.0275, it can be concluded intuitively that the best value of a for image coding is somewhere in the mid-range [2].

The dual pair filter coefficients for the decomposition and reconstruction filters are tabulated in Tables 3.8, 3.9, 3.10 and 3.11 for $a = 0.03125$, $a = 0.05000$, $a = 0.06250$ and $a = 0.07500$ respectively.

n	$h(n)$	$g(n)$	$\tilde{h}(n)$	$\tilde{g}(n)$
-4	0	0.009375	0	0
-3	0	-0.075000	-0.009375	-0.031250
-2	-0.031250	-0.259375	-0.075000	-0.250000
-1	0.250000	0.650000	0.259375	0.562500
0	0.562500	-0.259375	0.650000	-0.250000
1	0.250000	-0.075000	0.259375	-0.031250
2	-0.031250	0.009375	-0.075000	0
3	0	0	-0.009375	0
4	0	0	0	0

Table 3.8: The Laplacian filter coefficients $h(n)$, $g(n)$, $\tilde{h}(n)$ and $\tilde{g}(n)$ for $a = 0.03125$.

n	$h(n)$	$g(n)$	$\tilde{h}(n)$	$\tilde{g}(n)$
-4	0	0.010714	0	0
-3	0	-0.053571	-0.010714	-0.050000
-2	-0.050000	-0.260714	-0.053571	-0.250000
-1	0.250000	0.607143	0.260714	0.600000
0	0.600000	-0.260714	0.607143	-0.250000
1	0.250000	-0.053571	0.260714	-0.050000
2	-0.050000	0.010714	-0.053571	0
3	0	0	-0.010714	0
4	0	0	0	0

Table 3.9: The Laplacian filter coefficients $h(n)$, $g(n)$, $\tilde{h}(n)$ and $\tilde{g}(n)$ for $a = 0.05000$.

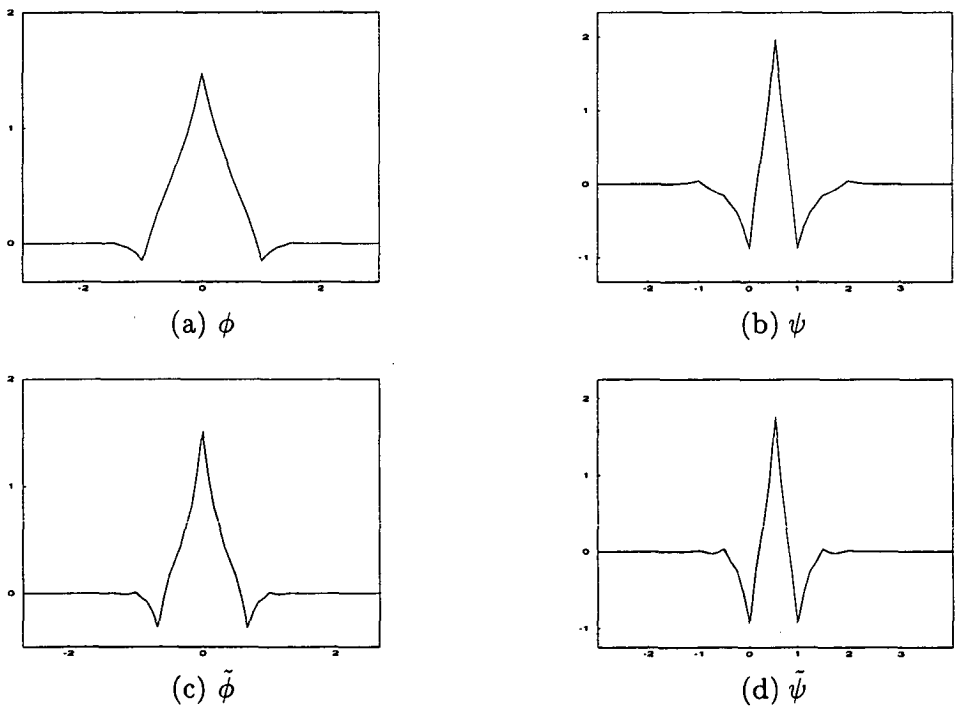
n	$h(n)$	$g(n)$	$\tilde{h}(n)$	$\tilde{g}(n)$
-4	0	0.010417	0	0
-3	0	-0.041667	-0.041667	-0.062500
-2	-0.062500	-0.260417	-0.041667	-0.250000
-1	0.250000	0.583333	0.260417	0.625000
0	0.625000	-0.260417	0.583333	-0.250000
1	0.250000	-0.041667	0.260417	-0.062500
2	-0.062500	0.010417	-0.041667	0
3	0	0	-0.041667	0
4	0	0	0	0

Table 3.10: The Laplacian filter coefficients $h(n)$, $g(n)$, $\tilde{h}(n)$ and $\tilde{g}(n)$ for $a = 0.06250$.

n	$h(n)$	$g(n)$	$\tilde{h}(n)$	$\tilde{g}(n)$
-4	0	0.009375	0	0
-3	0	-0.031250	-0.009375	-0.075000
-2	-0.075000	-0.259375	-0.031250	-0.250000
-1	0.250000	0.562500	0.259375	0.650000
0	0.650000	-0.259375	0.562500	-0.250000
1	0.250000	-0.031250	0.259375	-0.075000
2	-0.075000	0.009375	-0.031250	0
3	0	0	-0.009375	0
4	0	0	0	0

Table 3.11: The Laplacian filter coefficients $h(n)$, $g(n)$, $\tilde{h}(n)$ and $\tilde{g}(n)$ for $a = 0.07500$.

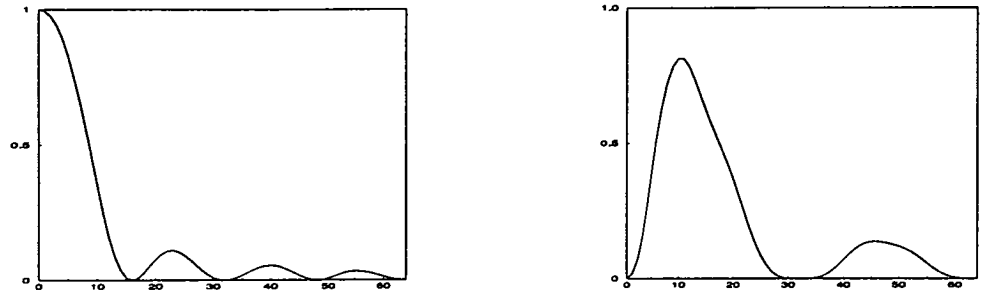
Figures 3.20, 3.21, 3.22, 3.23, 3.24, 3.25, 3.26 and 3.27 show the scaling and wavelet functions for analysis and synthesis and their spectra for $a = 0.03125$, $a = 0.05000$, $a = 0.06250$ and $a = 0.07500$ respectively. It can be seen that for the $a = 0.05000$, both ϕ and $\tilde{\phi}$ are very close to each other and very close to a coiflet for $L = 2$.



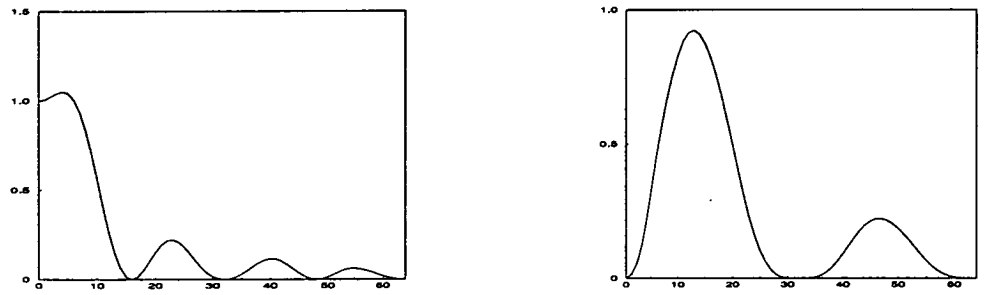
(a) ϕ (b) ψ

(c) $\tilde{\phi}$ (d) $\tilde{\psi}$

Figure 3.20: The biorthonormal close to orthonormal scaling and wavelet functions for $a = 0.03125$. (a) and (b) for decomposition, (c) and (d) for reconstruction.

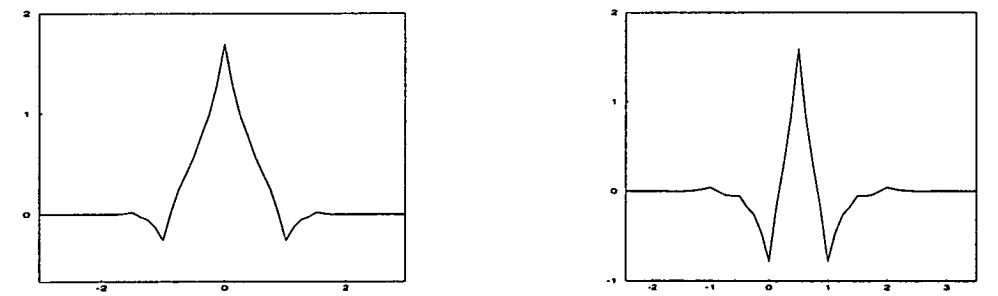


(a) Spectrum of ϕ (b) Spectrum of ψ

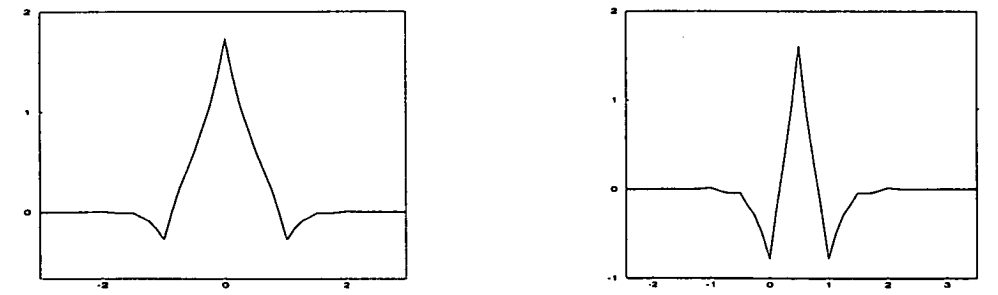


(c) Spectrum of $\tilde{\phi}$ (d) Spectrum of $\tilde{\psi}$

Figure 3.21: The spectra of the biorthonormal close to orthonormal scaling and wavelet functions for $a = 0.03125$. (a) and (b) for decomposition, (c) and (d) for reconstruction.

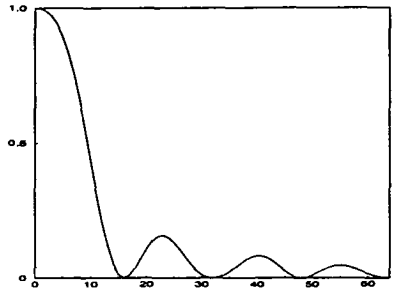


(a) ϕ (b) ψ

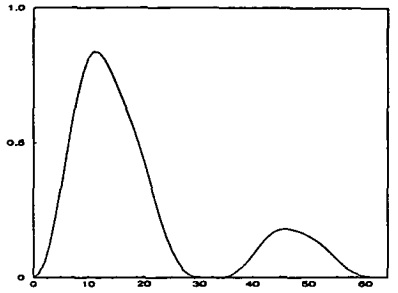


(c) $\tilde{\phi}$ (d) $\tilde{\psi}$

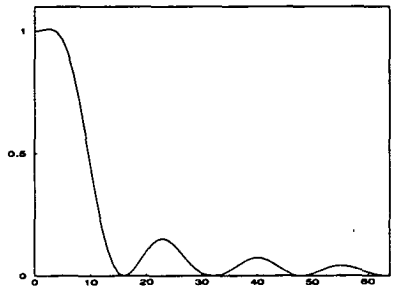
Figure 3.22: The biorthonormal close to orthonormal scaling and wavelet functions for $a = 0.05000$. (a) and (b) for decomposition, (c) and (d) for reconstruction.



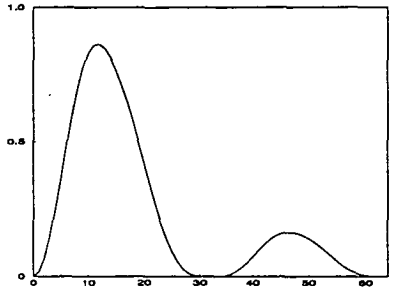
(a) Spectrum of ϕ



(b) Spectrum of ψ

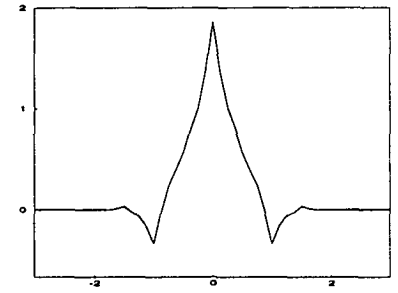


(c) Spectrum of $\tilde{\phi}$

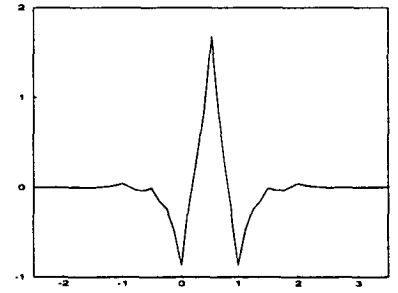


(d) Spectrum of $\tilde{\psi}$

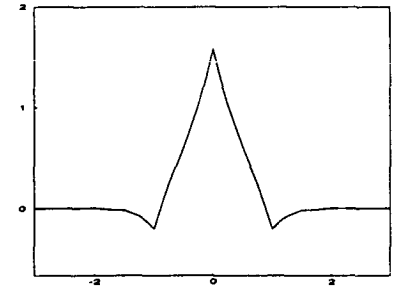
Figure 3.23: The spectra of biorthonormal close to orthonormal scaling and wavelet functions for $a = 0.05000$. (a) and (b) for decomposition, (c) and (d) for reconstruction.



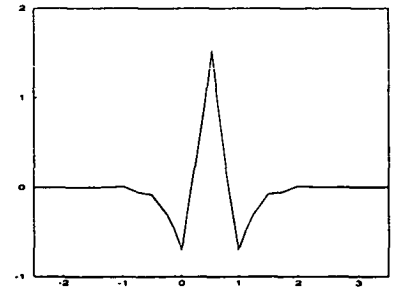
(a) ϕ



(b) ψ

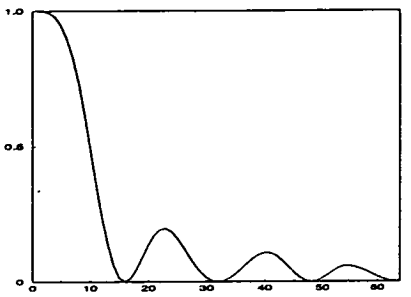


(c) $\tilde{\phi}$

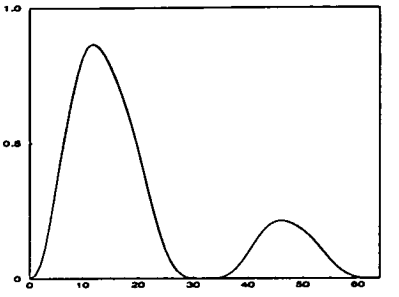


(d) $\tilde{\psi}$

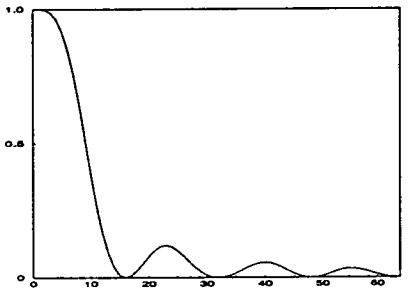
Figure 3.24: The biorthonormal close to orthonormal scaling and wavelet functions for $a = 0.06250$. (a) and (b) for decomposition, (c) and (d) for reconstruction.



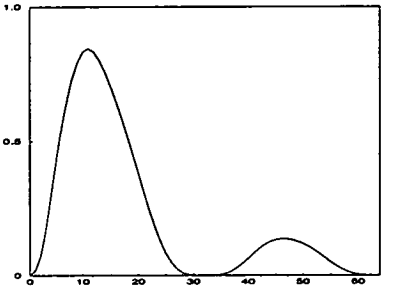
(a) Spectrum of ϕ



(b) Spectrum of ψ

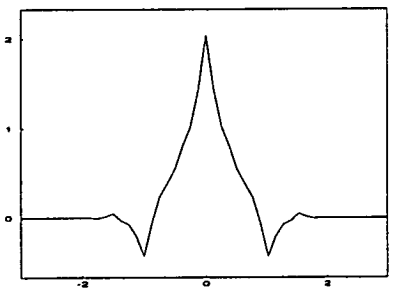


(c) Spectrum of $\tilde{\phi}$

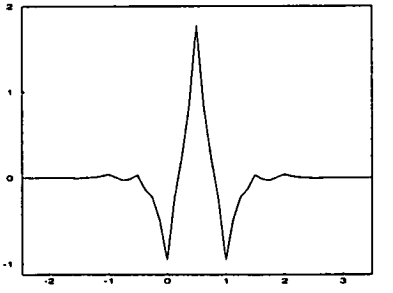


(d) Spectrum of $\tilde{\psi}$

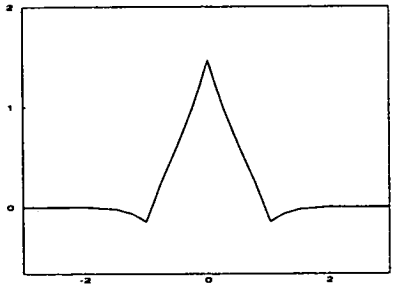
Figure 3.25: The spectra of biorthonormal close to orthonormal scaling and wavelet functions for $a = 0.06250$. (a) and (b) for decomposition, (c) and (d) for reconstruction.



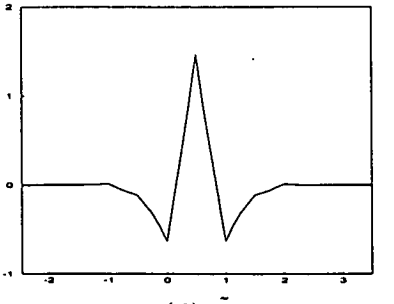
(a) ϕ



(b) ψ

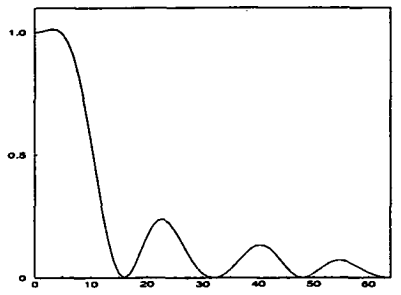


(c) $\tilde{\phi}$

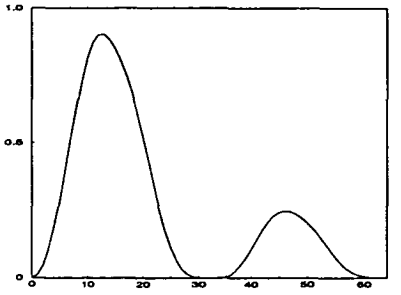


(d) $\tilde{\psi}$

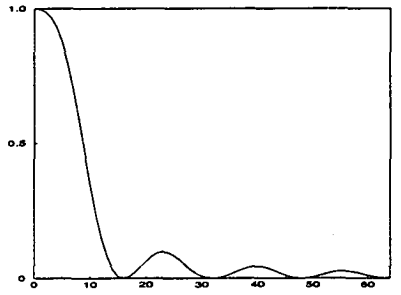
Figure 3.26: The biorthonormal close to orthonormal scaling and wavelet functions for $a = 0.07500$. (a) and (b) for decomposition, (c) and (d) for reconstruction.



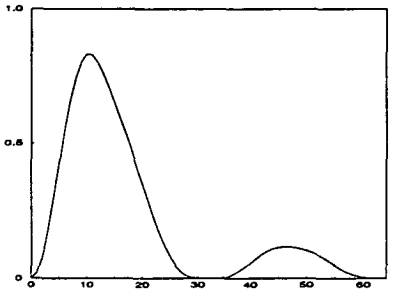
(a) Spectrum of ϕ



(b) Spectrum of ψ



(c) Spectrum of $\tilde{\phi}$



(d) Spectrum of $\tilde{\psi}$

Figure 3.27: The spectra of biorthonormal close to orthonormal scaling and wavelet functions for $a = 0.07500$. (a) and (b) for decomposition, (c) and (d) for reconstruction.

3.7 Summary

The family of orthonormal wavelets can be divided into two families, both of them compactly supported, i.e.

- One family is constructed such that for each L there is a unique solution and a maximum number of vanishing moments for the wavelet function only. This family is called the Daubechies wavelets. They are supported on an interval of length $2L - 1$. Their regularity increases linearly with L and is approximately equal to $0.2075L$ for large L. The wavelets with high regularity are included in this family.
- The other family is constructed such that the wavelet bases have a maximum number of vanishing moments for both the scaling and wavelet functions, i.e. coiflets. They are supported on an interval with length $3L - 1$.

The construction of biorthonormal wavelets results in symmetrical wavelet bases. The scaling functions of biorthonormal wavelets are B-splines, and the wavelets are also spline functions. The spline bases have the advantages that they have linear phase and are compactly supported. All the filter coefficients are dyadic rationals and this make them very suitable for fast computation. However, for small filter lengths, the dual functions have very low regularity. The biorthonormal wavelet bases can also be constructed to be close to an orthonormal wavelet by using the appropriate choice of $\mathcal{R}$ function, i.e. the Laplacian filter.

It can be concluded that the construction of both orthonormal and biorthonormal wavelet bases has several properties such as :

- Compact support.
- Smoothness.
- Vanishing moments.
- Asymmetry or Symmetry.

The design of wavelets with these properties very much depends on their intended applications.

Chapter 4

Psychovisual Coding of Wavelet Transform Coefficients

4.1 Introduction

One of the properties of the wavelet transform is that its basis functions offer excellent localisation properties in both the spatial and frequency domains and can therefore retain the details and sharp edges of an image better than conventional Fourier-based transforms. The wavelet transform decomposes the image's spectrum into a bank of bandpass subbands on a logarithmic scale, in contrast to the linear frequency bands of the Fourier transform [18] [72] [98].

The application of wavelet transforms in signal processing can be divided into two approaches, namely the conventional approach and the best basis approach. In the conventional approach, the wavelet function is recursively applied to a coarser approximation signal, which is obtained from the scaling function at each resolution, to extract the difference in information between consecutive resolutions. This scheme is also known as Mallat's pyramid scheme [18]. The best basis approach, which was introduced by Coifman, Meyer, Quaker and Wickerhauser [110] [111], involves transforming the signal into a set of over-complete basis functions, and permits the choice of a basis function which best suits the signals.

This chapter is divided into two sections. The first section discusses the conventional wavelet transform coding scheme using psychovisual coding of the wavelet transform coefficients. Psychovisual coding is then compared with statistical coding using the various wavelet bases discussed in the previous chapter. The last part of this chapter presents the best basis wavelet transform coding scheme.

4.2 Image representation of Mallat's Pyramid scheme

As stated in Section 3.5 the multiresolution decomposition of a signal can be obtained by computing a coarser approximation from a finer one, as well as from the difference in information between the two successive approximations. Since an image is two-dimensional, this section extends the multiresolution decomposition of a signal in one-dimension into two-dimensions.

A multiresolution decomposition of an image $I(x,y)$ of $\mathbf{L}^2(\mathbf{R}^2)$ is a sequence of subspaces of $\mathbf{L}^2(\mathbf{R})$ which satisfy a two-dimensional extension of the one-dimensional multiresolution properties. Let $\mathbf{V}_j$ for $j \in \mathbf{Z}$ be a multiresolution representation of $\mathbf{L}^2(\mathbf{R}^2)$. The approximation of an image $I(x,y)$ at resolution 2^j is equal to its orthogonal projection on the vector space $\mathbf{V}_j$. From property number 4 of the multiresolution decomposition (in Section 3.4), there exists a unique scaling function $\Phi(x,y)$ whose dilations and translations form an orthonormal basis for each space $\mathbf{V}_j$. The functions $\Phi(x,y)$ constitute an orthonormal basis for $\mathbf{V}_j$, generated by the $\mathbf{Z}^2$-translation of a single function Φ. This implies that

$$\Phi(x,y) = 2^{-j}\Phi(2^{-j}x - n_1, 2^{-j}y - n_2) \qquad n_1, n_2 \in \mathbf{Z}^2. \tag{4.1}$$

For such a multiresolution approximation, each vector $\mathbf{V}_j$ can be decomposed as a tensor product of two identical subspaces V_{j-1}, of $\mathbf{L}^2(\mathbf{R})$ [103], i.e.

$$\mathbf{V}_{j-1} = V_{j-1} \otimes V_{j-1}. \tag{4.2}$$

The sequence of vector spaces $\mathbf{V}_{j-1}$ for $j \in \mathbf{Z}$, forms a multiresolution approximation of $\mathbf{L}^2(\mathbf{R}^2)$ if and only if V_{j-1} is a multiresolution approximation of $\mathbf{L}^2(\mathbf{R})$ [18]. This

implies that $\Phi(x,y)$ can be rewritten as

$$\Phi(x,y) = \phi(x)\phi(y) \tag{4.3}$$

where $\phi(x)$ and $\phi(y)$ are the one-dimensional scaling functions of V_{j-1} and the two variables x and y are dilated separately. The orthogonal basis of $\mathbf{V}_{j-1}$ is then given by

$$2^{-j}\Phi_j(2^{-j}x - n_1, 2^{-j}y - n_2) = (2^{-j}\phi_j(2^{-j}x - n_1)\phi_j(2^{-j}y - n_2)) \qquad n_1, n_2 \in \mathbf{Z}^2. \tag{4.4}$$

Therefore, the approximation of an image $I(x,y)$ at resolution 2^j is characterized by the set of inner products

$$I_j = \langle I(x,y), \phi_j(2^{-j}x - n_1)\phi_j(2^{-j}y - n_2)\rangle \qquad n_1, n_2 \in \mathbf{Z}^2 \tag{4.5}$$

where $\Phi_j(x,y) = 2^{-j}\phi(2^{-j}x, 2^{-j}y)$.

As in the one-dimensional case, the complement space $\mathbf{W}_j$ is the orthogonal complement in $\mathbf{V}_{j-1}$ of $\mathbf{V}_j$ for $j \in \mathbf{Z}$. An orthonormal basis of this orthogonal complement can be constructed by scaling and translating three wavelet functions, i.e.,

$$\begin{aligned} \Psi^1(x,y) &= \phi(x)\psi(y) \\ \Psi^2(x,y) &= \psi(x)\phi(y) \\ \Psi^3(x,y) &= \psi(x)\psi(y). \end{aligned} \tag{4.6}$$

The difference of information is given by the three detail images of $I(x,y)$, i.e.,

$$\begin{aligned} d_j^1 &= \langle I(x,y), \Psi_j^1(2^{-j}x - n_1, 2^{-j}y - n_2)\rangle \qquad n_1, n_2 \in \mathbf{Z}^2 \\ d_j^2 &= \langle I(x,y), \Psi_j^2(2^{-j}x - n_1, 2^{-j}y - n_2)\rangle \qquad n_1, n_2 \in \mathbf{Z}^2 \\ d_j^3 &= \langle I(x,y), \Psi_j^3(2^{-j}x - n_1, 2^{-j}y - n_2)\rangle \qquad n_1, n_2 \in \mathbf{Z}^2 \end{aligned} \tag{4.7}$$

where $\Psi_j(x,y) = 2^{-j}\Psi(2^{-j}x, 2^{-j}y)$.

Equations (4.5) and (4.7) show that in two dimensions the inner products which define I_j, d_j^1, d_j^2 and d_j^3 are equal to a uniform sampling of the two-dimensional convolution products. Since the three wavelets Ψ^1, Ψ^2 and Ψ^3 are given by separable products of

the functions ϕ and ψ, these convolutions can be written as

$$\begin{aligned} I_j &= ((I(x,y) \odot \phi_j(-2^{-j}x)\phi_j(-2^{-j}y))(n_1,n_2)) \qquad n_1,n_2 \in \mathbf{Z}^2 \\ d_j^1 &= ((I(x,y) \odot \phi_j(-2^{-j}x)\psi_j(-2^{-j}y))(n_1,n_2)) \qquad n_1,n_2 \in \mathbf{Z}^2 \\ d_j^2 &= ((I(x,y) \odot \psi_j(-2^{-j}x)\phi_j(-2^{-j}y))(n_1,n_2)) \qquad n_1,n_2 \in \mathbf{Z}^2 \\ d_j^3 &= ((I(x,y) \odot \psi_j(-2^{-j}x)\psi_j(-2^{-j}y))(n_1,n_2)) \qquad n_1,n_2 \in \mathbf{Z}^2. \end{aligned} \tag{4.8}$$

Equation (4.8) shows that in two dimensions the approximation images and detail images are computed by a separable filtering of the image along the vertical and horizontal directions.

Decomposition and Reconstruction of the Wavelet Representation for Images

In two dimensions, the wavelet transform can be computed by a separable extension of the one-dimensional decomposition algorithm [18]. At each step an image is decomposed into I_1, d_1^1, d_1^2 and d_1^3. This algorithm is illustrated by the block diagram in Figure 4.1. Firstly, the rows of an image I_0 are convolved with one-dimensional filters

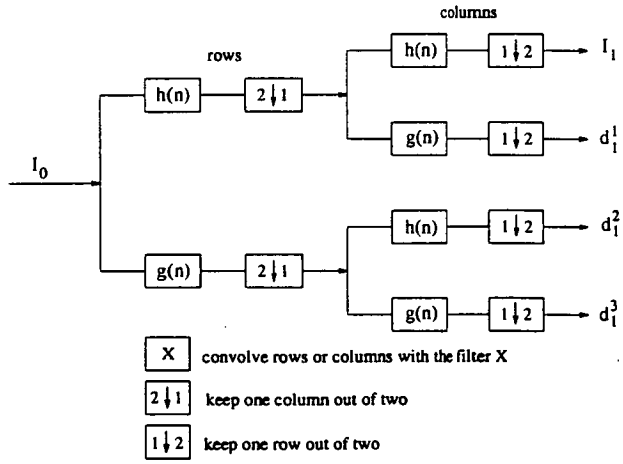


Figure 4.1: Two-dimensional decomposition of an image for $j = 1$ using the pyramid scheme.

H and G, and every other row is retained; then the columns are convolved with the filters H and G, and every other column is retained. The filters H and G are quadra-

ture mirror filters. The wavelet transform is then computed by repeating this process for $1 \leq j \leq J$. This scheme is called Mallat's fast wavelet transform decomposition.

The reconstruction algorithm for two dimensions can also be extended from the one-dimensional algorithm. Figure 4.2 shows the block diagram for the reconstruction of an image. An image I_0 is reconstructed from I_1, d_1^1, d_1^2 and d_1^3. Between each column

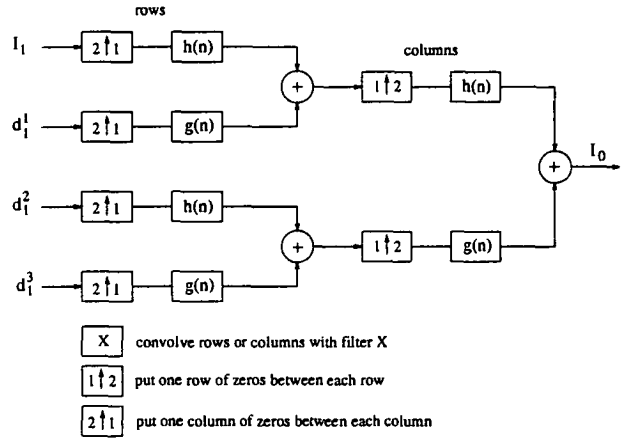


Figure 4.2: Two-dimensional reconstruction of an image I_{j-1} for $j = 1$ using the pyramid scheme.

of I_1, d_1^1, d_1^2 and d_1^3 a column of zeros is inserted, then the rows are convolved with one-dimensional filters, a row of zeros between each row of the resulting image is added, and the columns are convolved with one-dimensional filters. The filters used in the reconstruction are also quadrature mirror filters, $\bar{H}$ and $\bar{G}$. The filters $\bar{H}$ and $\bar{G}$ are conjugate filters of H and G respectively.

If the original image I_0 is an $N \times N$ array of pixels, then every array I_1, d_1^1, d_1^2 and d_1^3 consists of $\frac{N}{2} \times \frac{N}{2}$ elements and each is represented by an image of one quarter the size of the original. For convenience, the decomposition notation is changed from I_1, d_1^1, d_1^2 and d_1^3 to V_1^{HH}, W_1^{HG}, W_1^{GH} and W_1^{GG} respectively. This representation shows more clearly which filters act on which subspace. Figure 4.3 shows the decomposition and reconstruction of an image for $j = 3$.

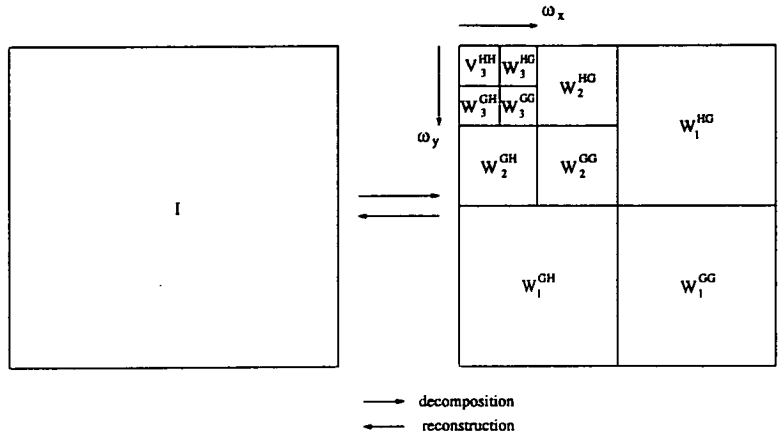


Figure 4.3: Decomposition and reconstruction of an image for 3 levels of decomposition ($j = 3$).

4.3 Conventional Wavelet Transform Coding Scheme

Psychovisual experiments of the HVS, discussed in Subsection 2.3.1, have produced sensitivity functions with a bandpass characteristic denoted by $S(\omega)$. $S(\omega)$ has a peak sensitivity between 3 to 3.5 cycles per degree. The sensitivity is reduced slightly at low frequencies but drops-off very sharply at higher frequencies. Therefore, the low amplitude high frequency components of the decomposed image can be readily discarded. Furthermore it is well known that errors are less visible in bright and busy areas, e.g. areas with significant edges and discontinuities in the image. In these areas coarse quantization can be used to reduce the number of bits required per pixel. The main computational step in psychovisual image compression is the subjective thresholding and quantization of the transform coefficients, allocating the most bits to the coefficients that are not de-emphasized by the HVS.

4.3.1 Extension of the HVS Model into 2 Dimensions

Equation (2.19) stated the relationship between the sensitivity of the HVS to small changes in an image and the spatial frequency of the altered information. One particular relationship of this form, proposed by Ngan et al. [58] and expanded into

two-dimensions is given by the following equation

$$S(\omega_x, \omega_y) = (0.31 + 0.69(\omega_x, \omega_y))e^{-0.29(\omega_x, \omega_y)} \tag{4.9}$$

where (ω_x, ω_y) are the spatial frequencies for the x and y coordinates in cycles per degree (cpd). By using the International Viewing Standard (illustrated in Appendix B) [112], i.e., that the distance between the viewer and the screen is maintained at a standard of six to eight times the picture height, a viewing angle θ of approximately 8 degrees is subtended by the viewer. The horizontal and vertical spatial frequencies of the HVS ω_h and ω_v, can be expressed by

$$\omega_h = A_r \frac{Hor}{\theta} \quad \text{cpd} \tag{4.10}$$

$$\omega_v = \frac{Ver}{\theta} \quad \text{cpd} \tag{4.11}$$

where Hor and Ver are the number of horizontal and vertical samples in the image respectively and A_r is the aspect ratio, i.e. image width to image height of $4:3$. The spatial frequency corresponding to the image coordinates (x, y) is given by

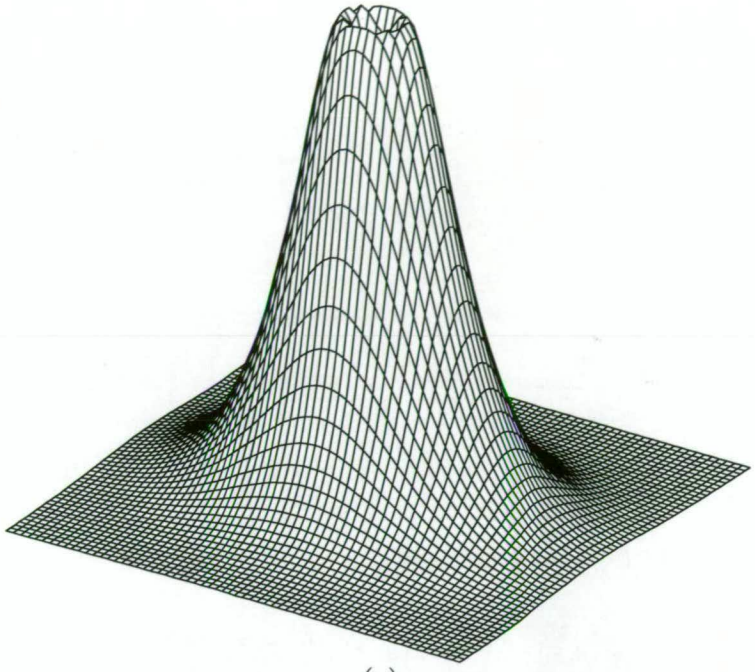
$$(\omega_x, \omega_y) = \sqrt{\left(\frac{\omega_v x}{N-1}\right)^2 + \left(\frac{\omega_h y}{N-1}\right)^2} \qquad \text{for} \quad x, y = 0, 1, ..., N-1 \tag{4.12}$$

where $N \times N$ is the image dimensions.

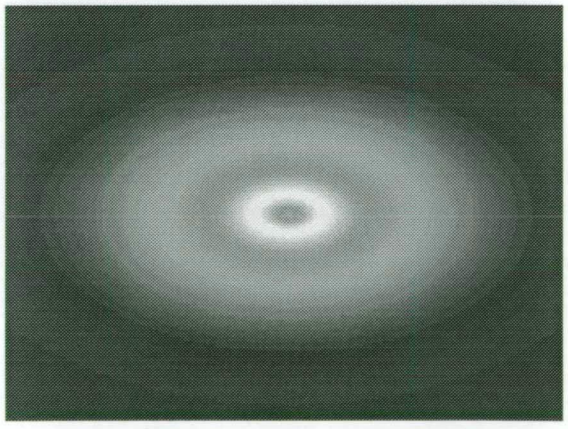
4.3.2 Subjective Thresholding

The sensitivity function shown in Figure 4.4 is obtained from psychovisual experiments involving a sinusoidal grating excitation and consequently cannot be used directly for thresholding the wavelet coefficients. Unlike the transform coefficients of the DCT, the wavelet transform coefficients do not represent a single discrete frequency. Instead each wavelet coefficient represents an octave subband with significant spectral overlap with neighbouring subbands. Therefore when we calculate the subband sensitivity for the wavelet transform coefficients at scale 2^j we must compute the HVS sensitivity function over the complete subband. The sensitivity factor for the coefficients at scale 2^j is calculated [2] [113] by

$$\hat{S}_j^k = k_e \left[\int \int |S(\omega_x, \omega_y)|^2 \left| \Psi_j^k(\omega_x, \omega_y) \right|^2 d\omega_x d\omega_y \right]^{\frac{1}{2}} \tag{4.13}$$



(a)



(b)

Figure 4.4: The spectrum of sensitivity function HVS. (a) Magnitude spectrum in 3D and (b) magnitude expressed as gray scale image.

where k_e is an energy constant, $S(\omega_x, \omega_y)$ is the HVS sensitivity function, $\Psi_j^k(\omega_x, \omega_y)$ is the wavelet frequency response of either $\Psi(x, y)$ or $\Phi(x, y)$ as stated in equation (4.6) or (4.3), and the superscripts $k = 1, 2, 3$correspond to HG, GH and GG respectively.

The simplest method for computing the subband sensitivity for the chosen wavelet at scale 2^j is to determine how that coefficient affects the spectrum of the image and then apply equation (4.13). This can be done by setting the appropriate coefficient to unit impulse in the subspace at scale 2^j and performing the inverse wavelet transform, followed by a Fourier transform [2] [113]. For example, a unit impulse in subband W_4^{GG} at the spatial coordinate $(47, 47)$ as shown in Figure 4.5 will produce the wavelet $\Psi_4(x, y)$ on application of an inverse wavelet transform. Application of the Fourier transform results in its spectrum, $\Psi_4^{GG}(\omega_x, \omega_y)$. Figure 4.6 (a) shows the corresponding wavelet function $\Psi_4^{GG}(x, y)$, and Figure 4.6 (b) shows the magnitude spectrum of $|\Psi_4^{GG}(\omega_x, \omega_y)|$.

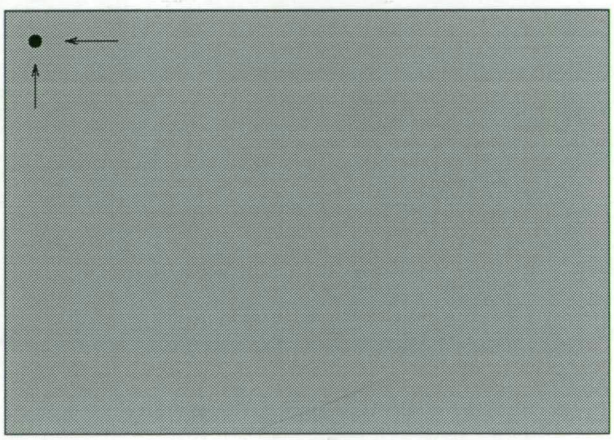
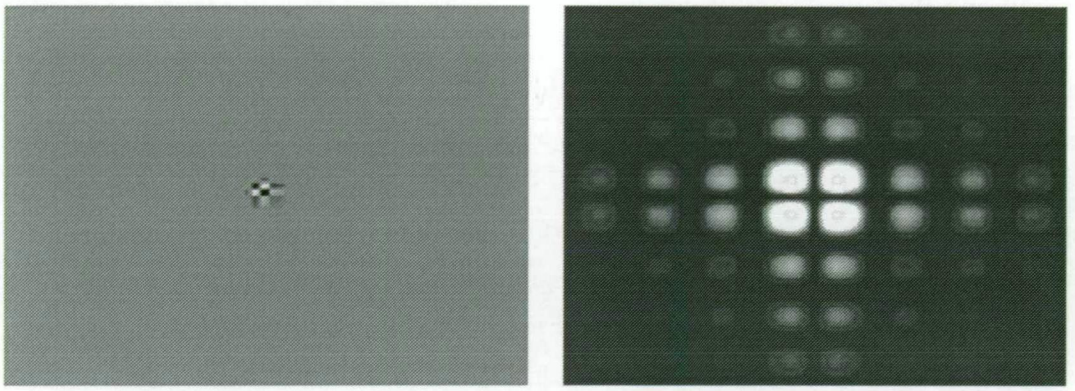


Figure 4.5: Impulse in subband W_4^{GG} at coordinate $(47, 47)$ in a 512 x 512 image plane.

The sensitivity factors for the constructed wavelet bases shown in Section 3.6 for each subband is shown in Tables 4.1, 4.2, 4.3, 4.4 and 4.5. The lowest subband sensitivity (V_4^{HH}) corresponding to the lowest frequency band is normalised to unity. In a multiresolution signal or image decomposition, the lowest subband contains a major percentage of the total energy present in the original signal or image. Therefore, the lowest subband is not subjectively thresholded. The lowest subband, however, is coded using a statistical scheme like that described in Section 2.2.



(a) (b)

Figure 4.6: Example of the Daubechies wavelet with $L = 2$ and its spectrum, expressed as a gray scale image. (a) the wavelet function, (b) the magnitude spectrum of the Fourier transform.

	$L = 2$	$L = 3$	$L = 4$	$L = 6$
subband	sensitivity	sensitivity	sensitivity	sensitivity
V_4^{HH}	1.000000	1.000000	1.000000	1.000000
W_4^{GH}	1.066019	1.087218	1.100307	1.116007
W_4^{HG}	1.071870	1.095126	1.109609	1.126880
W_4^{GG}	0.933435	0.977233	1.005256	1.039647
W_3^{GH}	0.683570	0.697494	0.706812	0.718035
W_3^{HG}	0.606645	0.617414	0.625066	0.634542
W_3^{GG}	0.443867	0.462818	0.476099	0.492854
W_2^{GH}	0.281039	0.278162	0.277953	0.278659
W_2^{HG}	0.210355	0.201992	0.198623	0.195978
W_2^{GG}	0.107184	0.104836	0.104731	0.105554
W_1^{GH}	0.068206	0.058771	0.054308	0.050160
W_1^{HG}	0.043194	0.033257	0.028573	0.024246
W_1^{GG}	0.011277	0.008472	0.007305	0.006317

Table 4.1: Sensitivity factors of orthonormal Daubechies wavelet bases.

	$L = 4$	$L = 6$
subband	sensitivity	sensitivity
V_4^{HH}	1.000000	1.000000
W_4^{GH}	1.100307	1.116007
W_4^{HG}	1.109608	1.126880
W_4^{GG}	1.005256	1.039647
W_3^{GH}	0.706812	0.718035
W_3^{HG}	0.625066	0.634542
W_3^{GG}	0.476099	0.492855
W_2^{GH}	0.277953	0.278659
W_2^{HG}	0.198623	0.195978
W_2^{GG}	0.104731	0.105554
W_1^{GH}	0.054308	0.050160
W_1^{HG}	0.028573	0.024246
W_1^{GG}	0.007305	0.006317

Table 4.2: Sensitivity factors of orthonormal near linear phase wavelet bases.

	$L = 2$	$L = 4$
subband	sensitivity	sensitivity
V_4^{HH}	1.000000	1.000000
W_4^{GH}	1.068043	1.102804
W_4^{HG}	1.074101	1.112379
W_4^{GG}	0.937515	1.010656
W_3^{GH}	0.684881	0.708617
W_3^{HG}	0.607608	0.626561
W_3^{GG}	0.445553	0.478703
W_2^{GH}	0.280650	0.277999
W_2^{HG}	0.209387	0.198083
W_2^{GG}	0.106840	0.104785
W_1^{GH}	0.067221	0.053569
W_1^{HG}	0.042218	0.027827
W_1^{GG}	0.010966	0.007122

Table 4.3: Sensitivity factors of orthonormal Coiflet wavelet bases.

	$\tilde{L}=2\ L=4$	$\tilde{L}+L=8$
subband	sensitivity	sensitivity
V_4^{HH}	1.000000	1.000000
W_4^{GH}	1.046088	1.124492
W_4^{HG}	1.026732	1.123249
W_4^{GG}	0.937998	1.029974
W_3^{GH}	0.573146	0.708672
W_3^{HG}	0.484981	0.611644
W_3^{GG}	0.354563	0.470357
W_2^{GH}	0.164594	0.269229
W_2^{HG}	0.115045	0.182653
W_2^{GG}	0.051042	0.096564
W_1^{GH}	0.024730	0.0485338
W_1^{HG}	0.015267	0.0239821
W_1^{GG}	0.002613	0.0057124

Table 4.4: Sensitivity factors of biorthonormal Spline wavelet bases.

	$a=0.03125$	$a=0.05000$	$a=0.06250$	$a=0.07500$
subband	sensitivity	sensitivity	sensitivity	sensitivity
V_4^{HH}	1.000000	1.000000	1.000000	1.000000
W_4^{GH}	1.103356	1.075661	1.063567	1.054261
W_4^{HG}	1.123558	1.082916	1.063444	1.047710
W_4^{GG}	0.987464	0.943667	0.928314	0.919110
W_3^{GH}	0.750007	0.695692	0.665711	0.639026
W_3^{HG}	0.677797	0.617158	0.584019	0.554976
W_3^{GG}	0.511176	0.452986	0.424466	0.401032
W_2^{GH}	0.339892	0.289183	0.261040	0.236277
W_2^{HG}	0.257462	0.214084	0.190708	0.170631
W_2^{GG}	0.141410	0.110043	0.094619	0.082101
W_1^{GH}	0.091282	0.069518	0.058367	0.049228
W_1^{HG}	0.057291	0.042827	0.035619	0.029851
W_1^{GG}	0.017223	0.011217	0.008600	0.006687

Table 4.5: Sensitivity factors of the biorthonormal close to the orthonormal wavelet bases.

After the sensitivity of a subband has been determined, a threshold can be applied to the wavelet coefficients of the corresponding subband. The subband threshold is defined [2] [113] by,

$$T_j^k = \frac{K}{\hat{S}_j^k}\sqrt{\frac{1}{N_s^2}\sum_{n_1}\sum_{n_2}\left|d_j^k(n_1,n_2)\right|^2} \tag{4.14}$$

where K is a proportionality constant determined from a subjective viewing of the quality of the reconstructed image and N_s is the subband dimension. The value of K is increased until the level at which a difference between the original image and reconstructed image is just noticeable by a human viewer (the subjective testing criterion is illustrated in Appendix B).

The square root of the summation function in equation (4.14) is the energy of the wavelet transform coefficients in that subband. The energy distribution for each of the subband transform coefficients using Daubechies wavelet bases with $L = 2$ for five standard images is shown in Figure 4.7.

It can be seen from these figures that the energy distributions for the standard images is relatively independent of the particular image being coded. This is true at least for typical images, but it is unlikely to hold for artificially-constructed images [54], for example an image consisting of only vertical or horizontal lines. Therefore, an average energy distribution is obtained using twenty typical images including the five standard images shown in Appendix A. The average energy distribution for these images is shown in Appendix C.

The optimum value for the constant K in equation (4.14) for various wavelet bases using this average energy distribution is shown in Table 4.6. These values are determined using the subjective testing criterion of Appendix B.

Once the original image is transformed into the wavelet transform basis set, the transformed images are then thresholded according to the subband thresholds given by equation (4.14). The wavelet transform coefficients, $X(n_1,n_2)$, are then thresholded using a standard thresholding method, given by

$$\mathcal{X}(n_1,n_2) = \begin{cases} X(n_1,n_2) & \text{if } |X(n_1,n_2)| \geq T_j^k \\ 0 & \text{if } |X(n_1,n_2)| < T_j^k \end{cases} \tag{4.15}$$

where $|X(n_1,n_2)|$ denote the absolute value of $X(n_1,n_2)$.

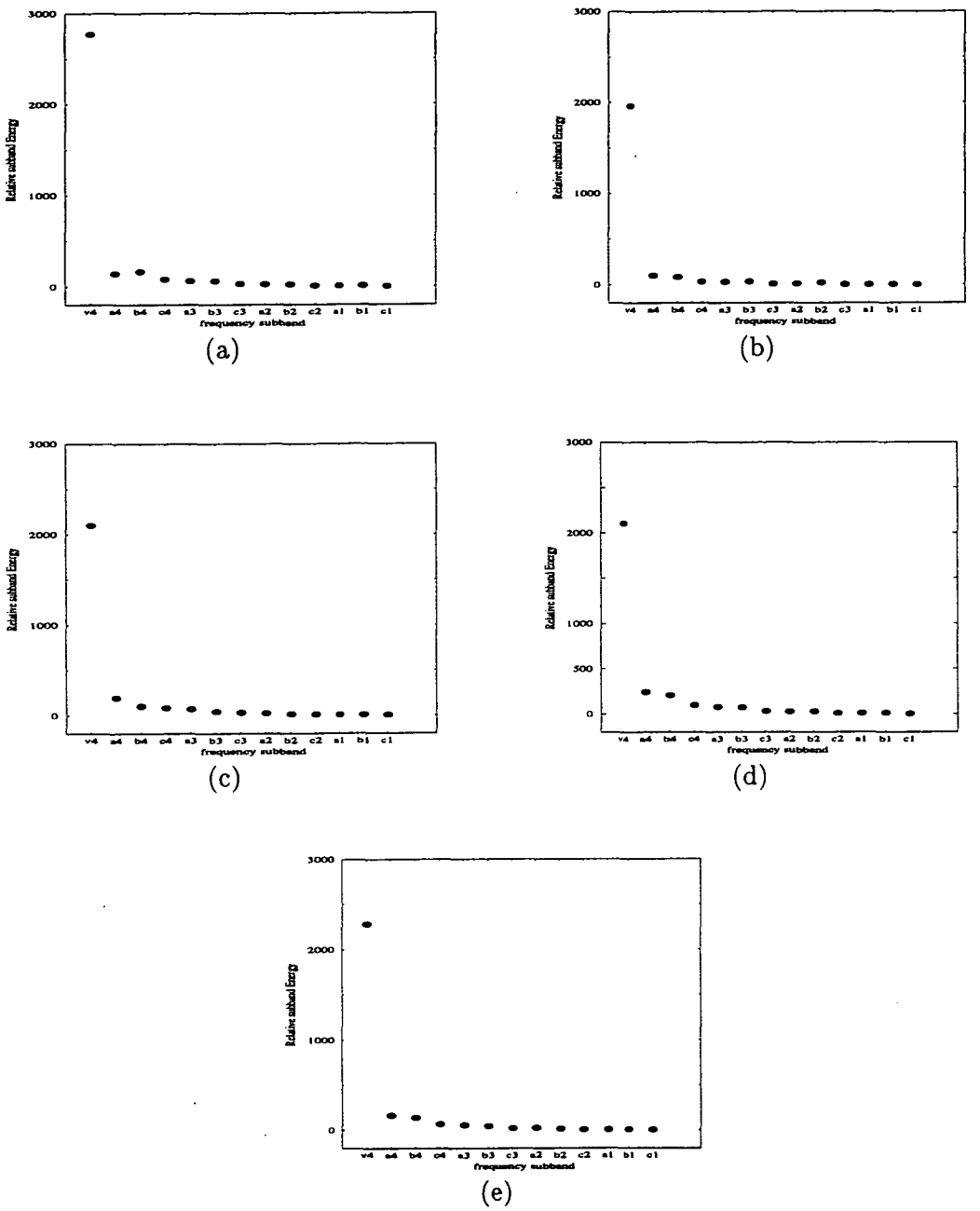


Figure 4.7: The subband energy distributions of the wavelet coefficients obtained using the Daubechies wavelet for $L = 2$, where on the horizontal axis v corresponds to V_4^{HH}, a corresponds to W^{GH}, b corresponds to W^{HG}, and c corresponds to W^{GG}. The numbers 1 to 4 correspond to the subband levels. (a) Test image "Airplane", (b) test image "Bird", (c) test image "Lenna", (d) test image "Peppers" and (e) test image "Zelda".

Type of wavelet	L	$\tilde{L}$	$L+\tilde{L}$	a	K
Daubechies	2	-	-	-	0.0500
	3	-	-	-	0.0434
	4	-	-	-	0.0408
	6	-	-	-	0.0387
Near linear	4	-	-	-	0.0404
phase	6	-	-	-	0.0381
Coiflet	2	-	-	-	0.0473
	4	-	-	-	0.0382
Spline	4	2	-	-	0.0248
	-	-	8	-	0.0358
Laplacian	2	2	-	0.03125	0.0651
	2	2	-	0.05000	0.0503
	2	2	-	0.06250	0.0433
	2	2	-	0.07500	0.0380

Table 4.6: The optimum value of K for various wavelet bases.

4.3.3 Subjective Quantization

The subjective quantization scheme is also based on the subband sensitivity of the transform coefficients. Once thresholded, the remaining coefficients are quantized into a number of discrete levels. Two quantization schemes are commonly used for the transform coefficients. These are Lloyd-Max quantization [114] [115] with fixed length code words, and uniform quantization with variable length code words [11, pp. 416-436]. The uniform quantization with variable length code words results in near optimum entropy coding [11, pp. 428-436]. This study uses the uniform quantization scheme. The non-zero transform coefficients, $X(n_1, n_2)$, are then quantized into $\overline{X}(n_1, n_2)$ according to

$$\overline{X}(n_1,n_2) = \begin{cases} \left[\frac{X(n_1,n_2)-T_j^k}{Q_j} + 0.50\right] & X(n_1,n_2) > 0 \\ \\ \left[\frac{X(n_1,n_2)+T_j^k}{Q_j} - 0.50\right] & X(n_1,n_2) < 0 \end{cases} \tag{4.16}$$

where $[\alpha]$ refers to the the integer value closet to α.

The subjective quantization also makes use of the effect of spatial masking allowing for a larger quantization step size in areas of high activity [11, pp. 416-436] [60]. The relationship between the quantization step sizes for sub-threshold distortion and the amount of activity in the subband has been the subject of previous research. This relationship is given by

$$Q_j^k = \sqrt{A_{F_j}^k T_j^k} \tag{4.17}$$

where $A_{F_j}^k$ is the activity function in the subband at scale 2^j, and $k = 1, 2, 3$ which correspond to the HG, GH and GG detail subbands respectively.

The activity function of each subband, A_F, defined as the sum of the horizontal and vertical gradients of luminance in a small area of an image, has been proposed in the past [11, pp. 416-436] [58] [60] [65] [116]. These activity functions have been calculated in both the spatial domain as well as from transform coefficients in the frequency domain. The most commonly proposed measure [60] is given by

$$A_F \propto \sum_{k=-1}^{+1} \sum_{l=-1}^{+1} (|M_{kl}^h| + |M_{kl}^v|) \tag{4.18}$$

where M_{kl}^h and M_{kl}^v are the horizontal and vertical slopes of the transform coefficients, given by

$$M_{kl}^h = X(n_1, n_2) - X(n_1 - 1, n_2)$$
$$M_{kl}^v = X(n_1, n_2) - X(n_1, n_2 - 1).$$

The application of equation (4.18), however, for a ramp function and saw-tooth function results in a number of inconsistencies, i.e. results in the same value of A_F. A more accurate measure of the activity function has been proposed [67] using the Laplacian edge detector. The Laplacian edge detector [117] achieves a second order differentiation through a spatial frequency approximation defined by

$$\omega^2 = (\omega_h^2 + \omega_h^2) \approx 4 - 2cos\omega_h - 2cos\omega_v \tag{4.19}$$

where ω_h and ω_v are the horizontal and vertical spatial frequencies corresponding to the transform coefficients position in a subband. In the z domain, this approximation

results in what is called a Laplacian mask and can be denoted by,

$$\mathcal{M}_L = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}. \tag{4.20}$$

Since video images have an aspect ratio 4 : 3, the mask is therefore altered to satisfy this aspect ratio. By inversely scaling the contributions of each of the transform coefficients to the mask output [67], the corresponding Laplacian mask can be obtained,

$$\mathcal{M}_L = \begin{bmatrix} -1.000 & -2.777 & -1.000 \\ -1.563 & 12.679 & -1.563 \\ -1.000 & -2.777 & -1.000 \end{bmatrix}. \tag{4.21}$$

The activity function is then given [67] by

$$A_{F_j}^k = 1 + q\sqrt{\frac{1}{N_s^2}\sum_k |\mathcal{M}_L \odot X(n_1, n_2)|^2} \tag{4.22}$$

where $\mathcal{M}_L \odot X(n_1, n_2)$ denotes the two-dimensional convolution output of the edge detector with $X(n_1, n_2)$, the wavelet transform coefficients at the subband scale 2^j, and q is a normalization factor derived from subjectively viewing reconstructed images. The optimum value for the normalization factor, q in equation (4.22) for various wavelet bases is shown in Table 4.7. The normalization factor, q, is subjectively adjusted so that the quantization of the subbands of coefficients results in only subthreshold distortion. The subjective testing criterion is illustrated in Appendix B.

4.3.4 Entropy Coding

After the thresholding and quantizing processes have been carried out, the remaining transform coefficients are then coded into binary streams for transmission or storage. In transform coding, the coefficients are scanned according to the frequency order in a particular sub-block, i.e. the coefficients are scanned from lower to higher coefficients by the zig-zag method illustrated in Figure 2.4. This method is very effective and useful, especially in low bit rate DCT coding [30], because high frequency components may be cut off when they are small in magnitude.

Type of wavelet	L	$\tilde{L}$	$L+\tilde{L}$	a	q
Daubechies	2	-	-	-	0.0135
	3	-	-	-	0.0152
	4	-	-	-	0.0163
	6	-	-	-	0.0168
Near linear	4	-	-	-	0.0150
phase	6	-	-	-	0.0158
Coiflet	2	-	-	-	0.0134
	4	-	-	-	0.0154
Spline	4	2	-	-	0.0161
	-	-	8	-	0.0167
Laplacian	2	2	-	0.03125	0.0130
	2	2	-	0.05000	0.0137
	2	2	-	0.06250	0.0144
	2	2	-	0.07500	0.0147

Table 4.7: The optimum value of q for various wavelet bases.

A property of the wavelet transform is that an image is decomposed into octave subbands, where the transformed image is arranged from a lower to higher frequency in the subband. As in the DCT, the wavelet transform coefficients are scanned in such a way as to improve the overall efficiency of the coding scheme. If we consider the decomposition of an image into the quad-tree structure depicted by Figure 4.3 (i.e. $3j+1$ subbands where $j=3$), then the scanning begins with V_3^{HH} then W_3^{GH}, W_3^{HG}, W_3^{GG} in that order, followed by W_2^{GH}, W_2^{HG}, W_2^{GG} and so on. This process is illustrated by Figure 4.8. In each subband the coefficients are coded using the same zig-zag scheme shown in Figure 2.4.

In each subband, the coefficients with non-zero magnitude are coded using their actual values and sequences of two or more zero values are coded by run-length codes which are equal to the number of successive zeros, preceded by a Run-Length Prefix (RLP). The RLP code is required in order to distinguish the run-length code from the amplitude code. The beginning of each new subband is marked by a Beginning Of Subbblock (BOS) marker, followed by the activity measure A_F for that subband, which is also required by the decoder. The coefficient magnitudes, the run-length, the BOS and the activity measure should all be represented by binary codes in order to produce the binary bit stream. This bit stream is then encoded using Huffman coding in which the statistical distribution of the stream is used to construct a look-up table of optimum

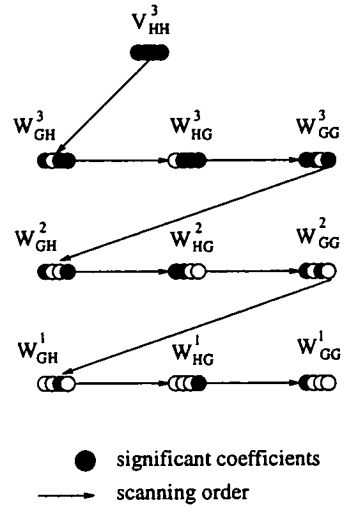


Figure 4.8: The scanning order for the wavelet coefficients.

variable length code-words. In practice, however, the exact statistical distribution of the stream is not often available. Fortunately the performance of Huffman coding is not particularly sensitive to small changes in distribution and therefore efficient code word assignment can be done using an average distribution of levels obtained from a number of typical images. In this study, the code-word tables are generated from the histograms averaged over the data produced by five typical images. The code-word tables are listed in Appendix D.

4.3.5 Compression Results

A comparison of the developed image compression scheme for different wavelet bases is presented in this section using standard test images. Since the evaluation of the quality of the reconstructed image is very difficult to judge objectively and consistently by a human observer, the performance criterion of the WPSNR (illustrated in Appendix E), with a constant target bit-rate of 0.40 bits per pixels (bpp) is used throughout this comparison. For some wavelet bases the bit rate of 0.40 bpp is not necessarily the optimum compression ratio, but in order to make a comparison of the reconstructed images either the bit rate or the WPSNR must be fixed. Since, it is very difficult to obtain the same WPSNR or image quality for different wavelet bases we have chosen to fix the bit rate.

In an attempt to illustrate quantitatively the benefits of psychovisual image compression, statistical image compression is also carried out. Both schemes employ very similar coding schemes, except that the statistical image compression uses uniform quantization. Figure 4.9 shows the block diagram of the statistical compression scheme and Figure 4.10 shows the block diagram of the psychovisual image compression technique [1] [2]. Table 4.8 shows the computed WPSNR for both schemes for each wavelet basis for constant bit rate of 0.40 for the standard image "Lenna". Table 4.9 shows the comparison of bit rates for statistical and psychovisual compression schemes for on the basis of comparable WPSNR. The benefit of psychovisual compression scheme is in the bit rate reduction (Table 4.9). Since the energy discarded from psychovisual thresholding is very small, the improve in WPSNR is small for same bit rates as expected (Table 4.8).

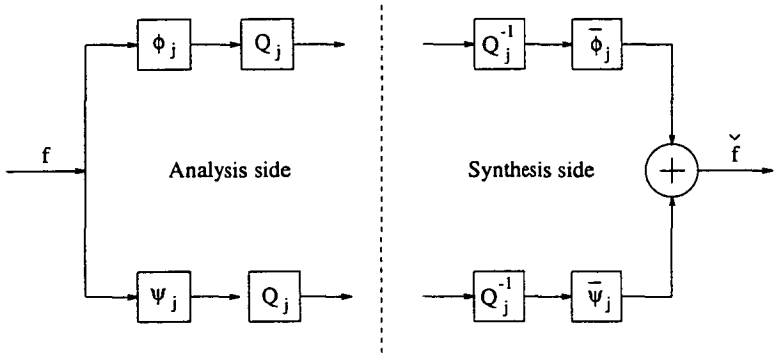


Figure 4.9: The block diagram of the statistical image compression scheme using the wavelet transform.

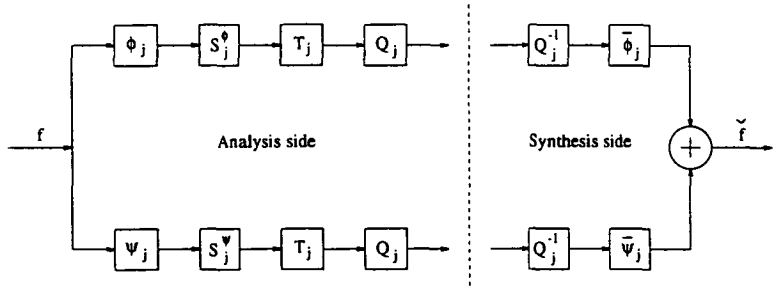


Figure 4.10: The block diagram of the psychovisual image compression scheme using the wavelet transform.

					Statistical	Psychovisual
Type of wavelet	L	$\tilde{L}$	$L+\tilde{L}$	a	WPSNR	WPSNR
Daubechies	2	-	-	-	37.609	38.125
	3	-	-	-	37.886	38.284
	4	-	-	-	37.944	38.301
	6	-	-	-	38.174	38.308
Near linear	4	-	-	-	38.468	38.656
phase	6	-	-	-	38.677	38.684
Coiflet	2	-	-	-	37.717	38.252
	4	-	-	-	38.494	38.734
Spline	4	2	-	-	38.363	40.196
	-	-	8	-	38.420	39.040
Laplacian	2	2	-	0.03125	36.613	36.677
	2	2	-	0.05000	37.522	38.145
	2	2	-	0.06250	37.760	38.528
	2	2	-	0.07500	38.035	39.018

Table 4.8: Comparison of results between the Statistical and Psychovisual compression schemes for constant bit rate of 0.40 bbp for the standard image "Lenna".

					Statistical		Psychovisual	
Type of wavelet	L	$\tilde{L}$	$L+\tilde{L}$	a	bit rate	WPSNR	bit rate	WPSNR
Daubechies	2	-	-	-	0.40	37.609	0.34	37.561
	3	-	-	-	0.40	37.886	0.33	37.827
	4	-	-	-	0.40	37.944	0.35	37.951
	6	-	-	-	0.40	38.174	0.36	38.105
Near linear	4	-	-	-	0.40	38.468	0.36	38.492
phase	6	-	-	-	0.40	38.677	0.37	38.737
Coiflet	2	-	-	-	0.40	37.717	0.34	37.873
	4	-	-	-	0.40	38.494	0.35	38.254
Spline	4	2	-	-	0.40	38.363	0.31	38.135
	-	-	8	-	0.40	38.420	0.35	38.243
Laplacian	2	2	-	0.03125	0.40	36.613	0.40	36.677
	2	2	-	0.05000	0.40	37.522	0.34	37.616
	2	2	-	0.06250	0.40	37.760	0.34	37.979
	2	2	-	0.07500	0.40	38.035	0.33	37.923

Table 4.9: Comparison of bit rates for Statistical and Psychovisual compression schemes for on the basis of comparable WPSNR.

It can be seen that psychovisual image compression results in a higher WPSNR compared to statistical image compression for each wavelet. This performance is not only in terms of a higher WPSNR but also in terms of the quality of the reconstructed image. In other words, if the quality of reconstructed image is constant (which, of course, in real terms cannot be easily achieved for different images) then the compression ratio achieved using psychovisual compression is higher than that achieved using statistical compression. Figures 4.11 and 4.12 show the original image "Lenna" and reconstructed versions using psychovisual compression for various wavelet bases. Figures 4.11 and 4.12 may appear to readers not having any discernible differences for different choices of wavelet basis due to the poor rendering of a standard laser printer. However, the difference can be observed on the monitor display (20 inch Sun colour monitor with the brightness of the image adjusted so that any unused portion of the screen around the image has zero perceived brightness). It is also true that some observers are unable to distinguish the difference between the two reconstructed images for different wavelet basis such as shown in the simulation results on Table 4.11.

The simulation results also show that the biorthonormal wavelet bases result in a higher WPSNR compared to the orthonormal wavelet bases for the same number of vanishing moments (the same filter length), with the exception of the Laplacian wavelet bases with $a < 0.05000$. The highest WPSNR is obtained using a B-spline wavelet with $L = 4$ and $\tilde{L} = 2$. For the orthonormal wavelet bases, increasing L results in a marginal increase in WPSNR. This means a slightly higher WPSNR can be obtained by using smoother wavelet bases at, however, a cost in computation. The invariance of the WPSNR to the filter length or smoothness of the orthonormal wavelets is attributable to the use of the HVS which de-emphasizes the high frequency ripples that result from the use of less regular wavelets. Moreover, although the sensitivity factors for orthonormal Daubechies wavelet bases, near linear phase wavelet bases and coiflet wavelet bases is similar for the same L values, the WPSNR obtained using coiflet wavelet bases is highest. These sensitivity factors are less for the higher subbands and more for the lower subbands.

The same trend in sensitivity factors can also be found for biorthonormal Spline wavelet bases. For the Laplacian wavelet bases, the sensitivity factors depend on the values of a. These sensitivity factors decrease for all subbands as the value of a increases. It is clear that the smoother the wavelet the less the sensitivity factor is in the lower subbands and the higher the sensitivity factors in the higher subbands. In subjective

(a)

(b)

(c)

Figure 4.11: Performance of psychovisual image compression. (a) Original image "Lenna", (b) reconstructed image using the orthonormal Daubechies wavelet basis and (c) reconstructed image using the orthonormal near linear phase wavelet basis.

(a)

(b)

(c)

Figure 4.12: Performance of psychovisual image compression. (a) Reconstructed image "Lenna" using the orthonormal coiflet wavelet basis, (b) reconstructed image "Lenna" using the biorthonormal spline wavelet basis and (c) reconstructed image "Lenna" using the biorthonormal Laplacian wavelet basis.

	As in Table 4.8	Swapping the wavelet
Type of wavelet	WPSNR [dB]	WPSNR [dB]
Spline		
$L = 4$ and $\tilde{L} = 2$	40.196	34.485
$L + \tilde{L} = 8$	39.040	38.892
Laplacian		
$a = 0.03125$	36.613	38.921
$a = 0.05000$	38.145	38.052
$a = 0.06250$	38.528	37.187
$a = 0.07500$	39.018	36.402

Table 4.10: Effect on the coding performance by interchanging analysis and synthesis wavelets.

thresholding, it is desirable to have higher sensitivity factors in the higher subbands, and lesser sensitivity factors in the lower subbands. In a multiresolution decomposition, the lower image subbands contain more important information than the higher subbands and therefore the transform coefficients in the higher subbands on application of the subjective thresholding can be discarded or coarsely quantized, resulting in fewer bits. The transform coefficients in the lower subbands are thresholded at a lower level due to the sensitivity of these subbands.

The psychovisual image compression, the subjective thresholding and quantization are only applied on the analysis side as illustrated in Figure 4.10. Therefore, only the analysis side of the wavelet bases is taken into account in the HVS system. In other words, only the analysis side plays a role in the subjective compression scheme. Consequently, this implies that when using biorthonormal wavelets less smooth wavelet bases or shorter filters should be used in the analysis side [1] [2]. This can be verified by swapping the decomposition and reconstruction filters and is demonstrated by the results in Table 4.10.

It can be seen from Table 4.10 that for biorthonormal spline wavelet bases, the WPSNRs with swapped filters (analysis filter swapped with synthesis filter) are less than compared to the WPSNRs without swapping. For biorthonormal Laplacian wavelet bases, the WPSNR with swapped filters for $a < 0.05000$ is higher compared to the WPSNR without swapping the filters and inversely for $a > 0.05000$. For $a = 0.05000$, however, the WPSNR with and without swapping filters is almost the same. This is

because for $a = 0.05000$, the analysis and the synthesis wavelet bases are very similar. The decrease in WPSNR for the B-spline orthonormal wavelets for $L = 4$ and $\tilde{L} = 2$ when swapping the filters demonstrates the importance of retaining the shorter filter on the analysis side. The less smooth wavelet bases result in more higher frequency ripples in their spectra, but if combined with the subjective thresholding scheme which occurs on the analysis side, then the higher frequency ripples are canceled out. In other words less regular wavelets result in lesser sensitivity factors. A smoother wavelet basis is required for the reconstruction side in order to recover a better quality reconstructed image, as pointed out in Table 4.8 for the orthonormal wavelet bases, i.e., the smoothest wavelet bases result in the highest WPSNR.

The performance of the final compression scheme developed in this chapter is also evaluated by subjective observation of the reconstructed image by thirty observers. The thirty observers consist of two observers with expertise in the field of image processing, five observers currently studying image processing at a postgraduate level, eighteen observers studying engineering at an undergraduate level and five observers in secondary school. Table 4.11 shows a comparison between the statistical and psychovisual coding schemes using this subjective assessment. Figure 4.13 shows a comparison between the statistical and psychovisual coding schemes for the test image "Lenna" using B-spline wavelet basis. The simulation results for the other test images using the psychovisual

Scheme	better	cannot distinguish	worse
Statistical	2	3	25
Psychovisual	25	3	2

Table 4.11: Results of the subjective assessment for statistical and psychovisual image compression using the standard image "Lenna".

image compression scheme are shown in Table 4.12. The original and the reconstructed images are also shown in Figures 4.14 and 4.15.

test image	compression ratio	WPSNR [dB]
"Airplane"	19.512	40.151
"Bird"	33.333	43.946
"Peppers"	21.053	39.200
"Zelda"	22.857	40.896

Table 4.12: Performance of psychovisual image compression for the test images "Airplane", "Bird", "Peppers" and "Zelda".

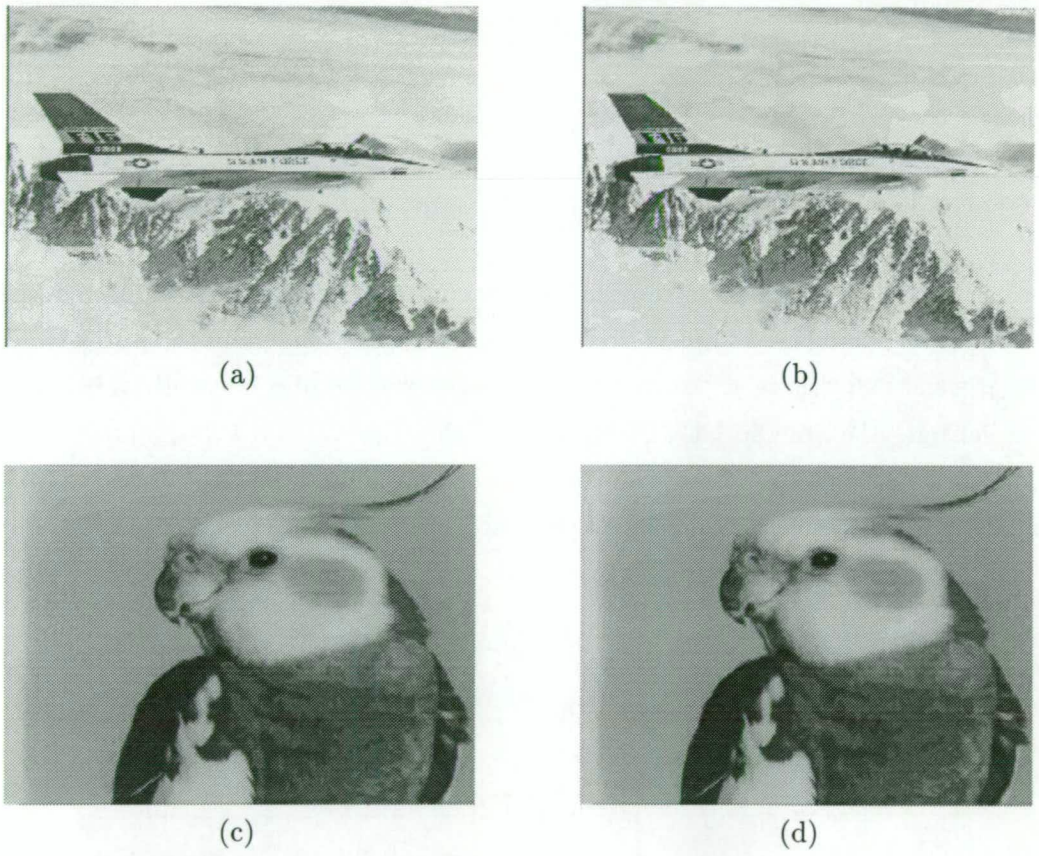
(a) (b)

Figure 4.13: A comparison of subjective assessment for the standard image "Lenna" for B-spline basis using (a) the psychovisual scheme and (b) the statistical scheme.

In order to see the benefits of psychovisual image compression scheme (proposed scheme), the simulation results are compared to the standard DCT scheme and the previous work by Ngan et al [58]. Table 4.13 shows the comparison of simulation results for the proposed scheme, the standard DCT scheme and the previous work by Ngan et al. It can be seen that the proposed scheme results in the highest PSNR compared to the standard DCT scheme and the scheme proposed by Ngan et al.

Type of scheme	test image	bit rate [bpp]	PSNR [dB]
Proposed scheme	"Lenna"	0.40	34.54
	"Peppers"	0.38	33.75
Standard DCT scheme	"Lenna"	0.42	32.60
	"Peppers"	0.40	31.38
Ngan et al [58]	"Lenna"	0.40	33.50
	"Peppers"	0.40	32.50

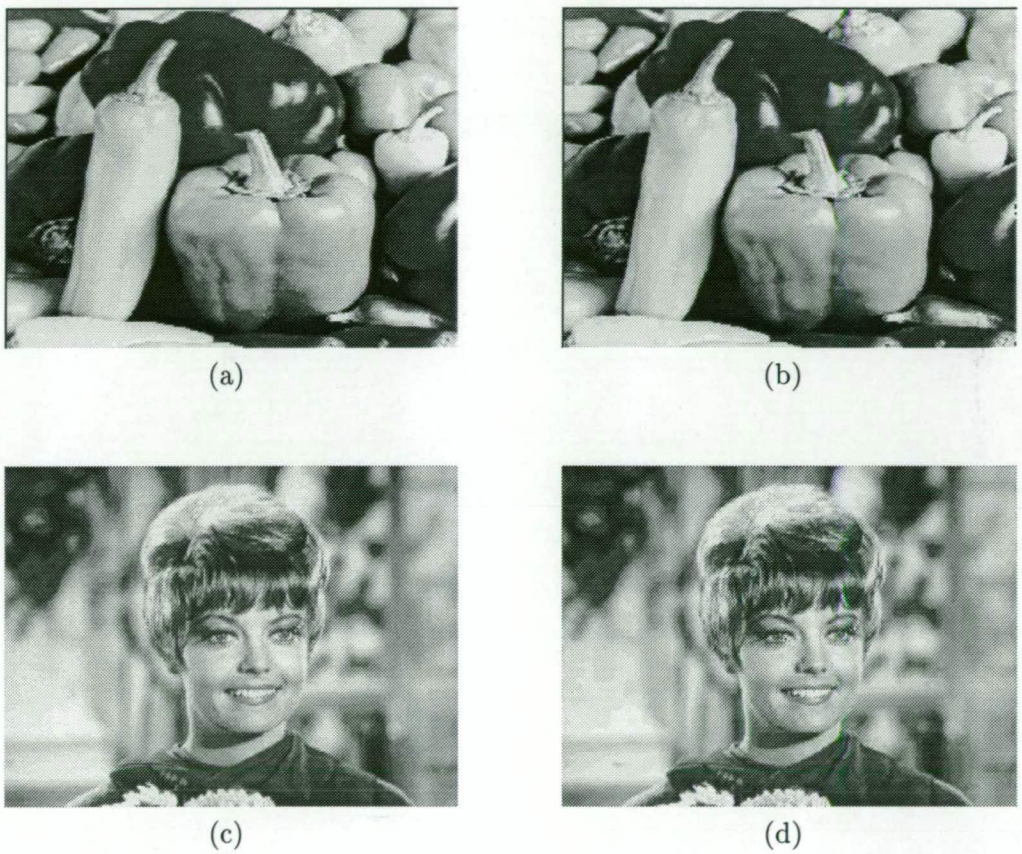
Table 4.13: Comparison of simulation results of three schemes.



(a) (b)

(c) (d)

Figure 4.14: Performance of psychovisual image compression, (a) The original and (b) the reconstructed test image "Airplane", (c) the original and (b) reconstructed test image "Bird".



(a) (b)

(c) (d)

Figure 4.14: Performance of psychovisual image compression. (a) The original and (b) the reconstructed test image "Peppers", (c) the original and (b) reconstructed test image "Zelda".

4.4 Best Basis Wavelet Transform Coding Scheme

In one-dimensional signal analysis the best basis wavelet scheme comprises the entire library of all permissible binary trees of maximum depth $log_2(N)$, growing out of the same root which is the original N-sample signal sequence. Figure 4.15 shows a complete binary tree of wavelet packet bases for the multiresolution decomposition of a one-dimensional signal. Figure 4.16 shows a complete quad-tree wavelet packet for the decomposition of an image. Each binary tree can be grown and pruned according to an algorithm based on some information cost function, so that the tree represents the best wavelet packet basis matched to the given signal or image.

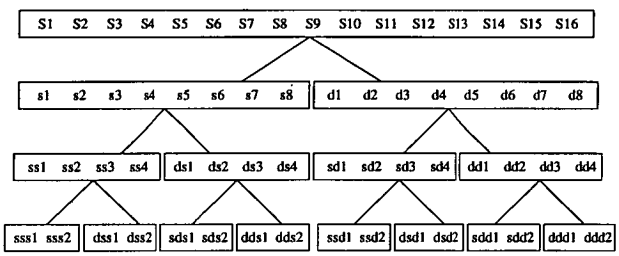


Figure 4.15: The complete wavelet basis set for the decomposition of a signal.

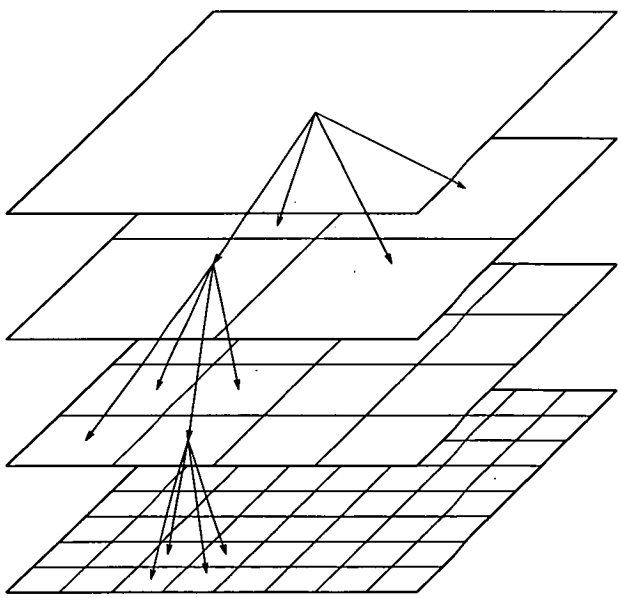
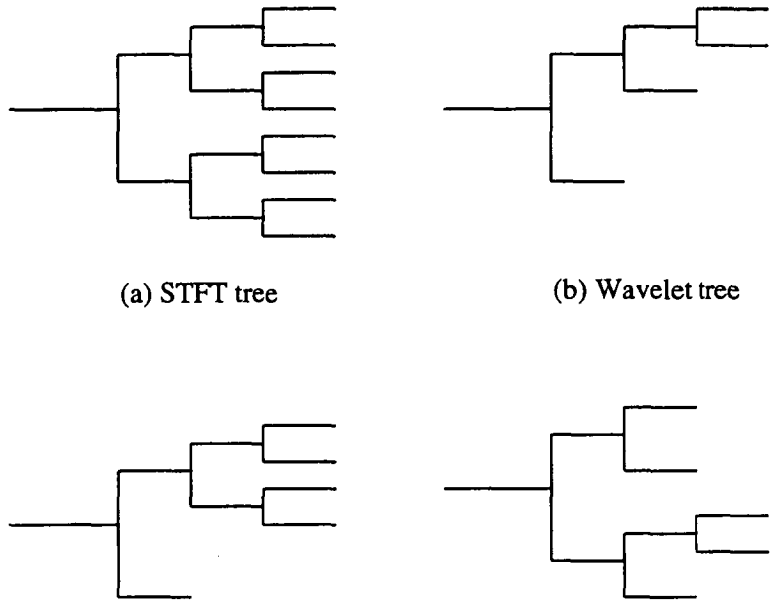


Figure 4.16: The complete wavelet basis set for the decomposition of an image.

It is therefore possible to use different wavelet packet bases for different segments of a non-stationary signal, or different areas of an image. It is also possible to use different

wavelet kernels for image segments of different characteristics. For example, a smooth section of an image may be more efficiently coded using a short-time Fourier transform (STFT) kernel rather than a compact but irregular wavelet kernel [119]. The STFT produces the most regular tree and the wavelet transform produces the most lopsided tree in which only the lowpass (leftmost) branches are allowed to have children. Figure 4.17 shows some permissible binary wavelet packets.



(a) STFT tree

(b) Wavelet tree

(c) Arbitrary best basis wavelet tree

Figure 4.17: Some permissible binary wavelet packets for 3 levels of decomposition.

In one-dimension, the conventional wavelet transform approach produces the most lopsided tree when only the lowpass (leftmost) branches are allowed to have children as illustrated in Figure 4.17 (b). In the best basis approach, the complete binary tree is grown and pruned according to an algorithm based on some information cost function, so that the final tree represents the wavelet packet basis best matched to the given signal.

The best basis wavelet approach is based on how to select the parent or its children depending on which is a more efficient encoding of the data. Each node represents the wavelet coefficients corresponding to a particular subspace in the wavelet decomposition of the signal. Once the original signal is transformed into an over complete basis set, an

information cost function is applied to the data in the new domain. The cost function is the entropy measure, which is defined by equation (2.5). The best basis for the signal can then be defined by extracting the set of subspaces with minimum information cost. The basis set that gives the most efficient representation for the signal is called the best basis. More details of wavelet packets can be found in [111] [112] [119] [120] [25] [121] [122].

4.4.1 Selecting the Best Basis

In two-dimensions, an image is decomposed into a quad tree set $\mathbf{W}$ of subspaces $W(j,l)$, where $j \geq 0$, and $0 \leq l < 4^j$. The 4 subspaces resulting from a single step of the decomposition are termed the children of the subspace, while the subspace itself is called the parent. Figure 4.18 shows an example of a complete basis for an image with three levels of decomposition.

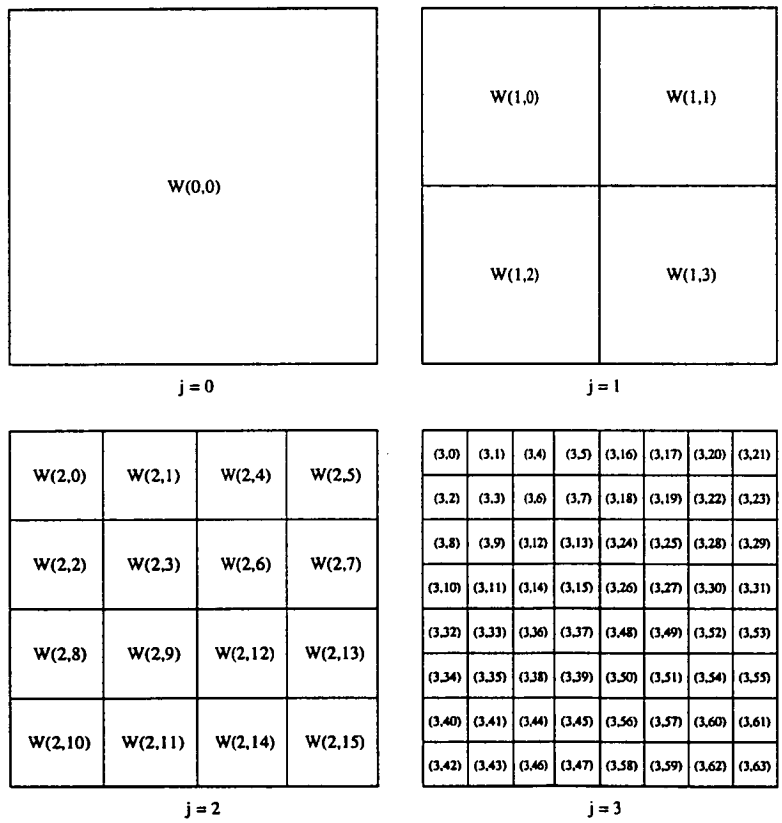


Figure 4.18: The complete basis of an image for three levels of decomposition.

The following steps are taken to select the best basis for an image:

1. Set a deepest level J which sets the maximum level of decomposition.
2. Label each subspace at level J as "kept" (the subspace at level J is indexed by (J,l) for $0 \leq l < 4^J$).
3. Compare the information cost of the subspace $W(j,l)$ with the sum of the information costs of its four children $W(j+1,4l), W(j+1,4l+1), W(j+1,4l+2)$ and $W(j+1,4l+3)$, starting with level $j = J-1$.
4. If the information cost of the parent is less than or equal to the sum of the children then mark the parent as "kept", and it is used to represent the image rather than the children. The tree is "pruned" or stops growing at this parent node. On the other hand, if the sum of the costs of the children is less than that of the parent, the parent is left unmarked but the recorded cost of transmission is the sum of the transmission costs of the children.
5. After all subspaces at level $j = J-1$ have been compared to their children, decrease the level index and repeat steps 4 and 5 above.
6. Continue in this way until the root $W(0,0)$ has been compared to its four children. The remaining "kept" nodes constitute the best basis, i.e. the basis with minimum information cost.

Before the selection of the best basis is carried out, the complete quad-tree of the transformed image is firstly thresholded according to the subband sensitivity of each subspace. In the best basis approach, the threshold function is taken into account in the same manner as in the conventional approach. The complete basis for the best basis approach, however, consists of the 4^j subspaces rather than $(3j+1)$ in the conventional approach. Therefore the threshold function is given [2] [113] by

$$T_j^l = \frac{K}{\hat{S}_j^l}\sqrt{\frac{1}{N_l^2}\sum_k \left|d_j^l(n_1,n_2)\right|^2} \tag{4.23}$$

where N_l is the subspace dimension and the subband sensitivity function $\hat{S}_j^l$ is computed [2] [113] by

$$\hat{S}_j^l = k_e\left[\int\int |S(\omega_x,\omega_y)|^2 \left|\Psi_j^l(\omega_x,\omega_y)\right|^2 d\omega_x d\omega_y\right]^{\frac{1}{2}} \tag{4.24}$$

where l is the subspace index ($0 \leq l < 4^j$).

The sensitivity of each subspace is then computed according to equation (4.23). If the original image has dimensions of 512 × 512 and three decomposition levels ($j = 3$), there will be 64 subspaces at level $j = 3$, 16 subspaces at level $j = 2$ and 4 subspaces at level $j = 1$, as illustrated in Figure 4.18. 16 subspaces of the second decomposition become parents of the 64 subspaces of the third decomposition. 4 subspaces of the first decomposition become parents of the 16 subspaces of the second decomposition. In order to calculate the sensitivity of each subspace for 3 levels of decomposition, the wavelet function is generated [3] as follows:

1. Set the subspace size to be 128 × 128.
2. Set an impulse at the coordinate (31,31), i.e. the center of the lowpass image and this impulse is inverse transformed to generate the wavelet function for the 128 × 128 subspace Φ^{HH}.
3. Similarly set an impulse at the coordinates (31,95), (95,31) and (95,95) and perform the inverse transform to generate the wavelet function for the 128 × 128 subspaces Ψ_1^{GH}, Ψ_1^{HG} and Ψ_1^{GG} respectively. We now have four subspaces in the top left subband corresponding to $j = 3$ at the top left of the quad tree.
4. Next, interpolate each subspace Φ_1^{HH}, Ψ_1^{GH}, Ψ_1^{HG} and Ψ_1^{GG} into a subspace with size 256 × 256 by filling the other 3 quadrants with zeros.
5. Then inverse transform each new signal subspace Φ_1^{HH}, Ψ_1^{GH}, Ψ_1^{HG} and Ψ_1^{GG} to generate the wavelet function for the 256 × 256 subspaces Φ_2^{HH}, Ψ_2^{GH}, Ψ_2^{HG} and Ψ_2^{GG} respectively.
6. Again interpolate each subspace Φ_2^{HH}, Ψ_2^{GH}, Ψ_2^{HG} and Ψ_2^{GG} of size 256 × 256 into the subspace of size 512 × 512.
7. Then inverse transform each new signal subspace (Φ_2^{HH}, Ψ_2^{GH}, Ψ_2^{HG} and Ψ_2^{GG}) to generate the wavelet function for the 512 × 512 subspaces Φ_3^{HH}, Ψ_3^{GH}, Ψ_3^{HG} and Ψ_3^{GG} respectively. Figure 4.19 illustrates this generation of wavelet functions.
8. Repeat steps 4 to 7 for the subbands in the top right, the bottom left and the bottom right at top left of the quad tree. Similarly perform the procedures for the top right, bottom left and bottom right of the quad tree.

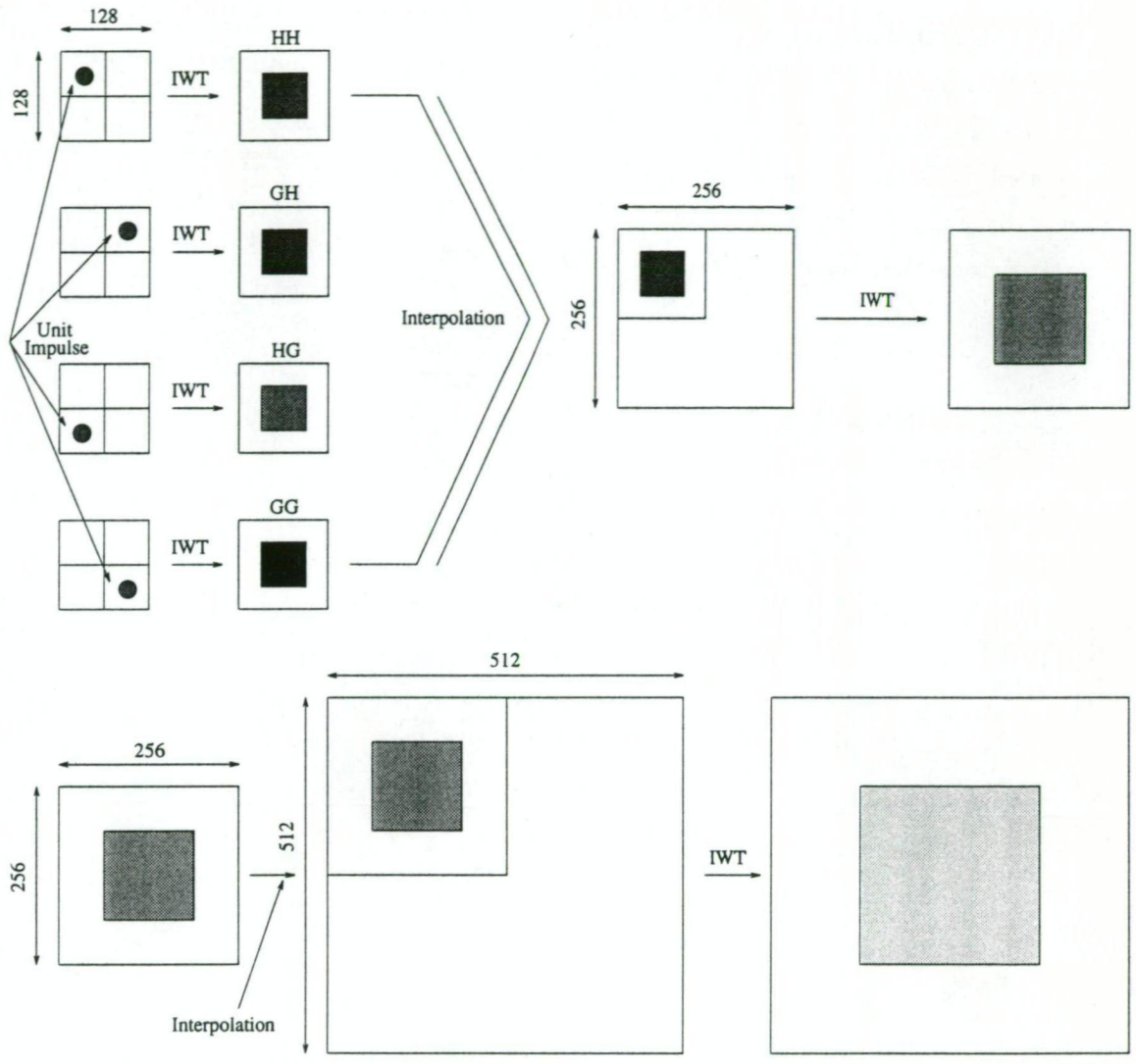


Figure 4.19: The diagram for generating the wavelet function for the best basis approach with 3 levels of decomposition. IWT is an inverse wavelet transform. HH, GH, HG and GG correspond to the wavelet functions of Φ^{HH}, Ψ^{GH}, Ψ^{HG} and Ψ^{GG} respectively.

To generate the wavelet functions of the second decomposition and the first decomposition, the above procedure is similarly carried out, with the beginning subspace size replaced by 256×256 and 512×512 respectively. The sensitivity factor of each subspace is calculated by taking the Fourier transform and using equation (4.24).

Since the biorthonormal spline wavelet basis for $L = 4$ and $\tilde{L} = 2$ resulted in the best compression ratio and the best reconstructed image using the conventional approach, only the results of the biorthonormal spline wavelet basis for $L = 4$ and $\tilde{L} = 2$ are presented for the best basis wavelet transform approach. The sensitivity factors of each subspace for 3 levels of decomposition are shown in Table 4.14, 4.15 and 4.16 for the first, second and third decompositions respectively.

subspace	sensitivity
$\hat{S}_1^0$	0.230073
$\hat{S}_1^1$	0.033369
$\hat{S}_1^2$	0.020600
$\hat{S}_1^3$	0.003526

Table 4.14: Sensitivity factors for the first decomposition.

subspace	sensitivity
$\hat{S}_2^0$	0.536165
$\hat{S}_2^1$	0.222094
$\hat{S}_2^2$	0.155236
$\hat{S}_2^3$	0.068873
$\hat{S}_2^4$	0.043846
$\hat{S}_2^5$	0.063087
$\hat{S}_2^6$	0.013633
$\hat{S}_2^7$	0.021288
$\hat{S}_2^8$	0.035408
$\hat{S}_2^9$	0.014980
$\hat{S}_2^{10}$	0.033754
$\hat{S}_2^{11}$	0.016170
$\hat{S}_2^{12}$	0.002940
$\hat{S}_2^{13}$	0.004461
$\hat{S}_2^{14}$	0.003355
$\hat{S}_2^{15}$	0.005286

Table 4.15: Sensitivity factors for the second decomposition.

subspace	sensitivity	subspace	sensitivity
$\hat{S}_3^0$	1.000000	$\hat{S}_3^{32}$	0.027762
$\hat{S}_3^1$	0.773373	$\hat{S}_3^{33}$	0.020334
$\hat{S}_3^2$	0.654407	$\hat{S}_3^{34}$	0.063007
$\hat{S}_3^3$	0.478428	$\hat{S}_3^{35}$	0.046448
$\hat{S}_3^4$	0.376629	$\hat{S}_3^{36}$	0.010139
$\hat{S}_3^5$	0.354726	$\hat{S}_3^{37}$	0.009354
$\hat{S}_3^6$	0.246815	$\hat{S}_3^{38}$	0.024737
$\hat{S}_3^7$	0.222398	$\hat{S}_3^{39}$	0.021818
$\hat{S}_3^8$	0.203480	$\hat{S}_3^{40}$	0.063363
$\hat{S}_3^9$	0.157126	$\hat{S}_3^{41}$	0.049477
$\hat{S}_3^{10}$	0.257360	$\hat{S}_3^{42}$	0.040244
$\hat{S}_3^{11}$	0.191317	$\hat{S}_3^{43}$	0.030299
$\hat{S}_3^{12}$	0.094736	$\hat{S}_3^{44}$	0.030794
$\hat{S}_3^{13}$	0.076971	$\hat{S}_3^{45}$	0.024499
$\hat{S}_3^{14}$	0.104335	$\hat{S}_3^{46}$	0.017107
$\hat{S}_3^{15}$	0.090588	$\hat{S}_3^{47}$	0.014515
$\hat{S}_3^{16}$	0.033671	$\hat{S}_3^{48}$	0.000981
$\hat{S}_3^{17}$	0.082693	$\hat{S}_3^{49}$	0.002199
$\hat{S}_3^{18}$	0.022925	$\hat{S}_3^{50}$	0.002381
$\hat{S}_3^{19}$	0.052224	$\hat{S}_3^{51}$	0.005163
$\hat{S}_3^{20}$	0.131512	$\hat{S}_3^{52}$	0.003581
$\hat{S}_3^{21}$	0.062261	$\hat{S}_3^{53}$	0.001665
$\hat{S}_3^{22}$	0.087931	$\hat{S}_3^{54}$	0.008902
$\hat{S}_3^{23}$	0.040070	$\hat{S}_3^{55}$	0.003993
$\hat{S}_3^{24}$	0.010793	$\hat{S}_3^{56}$	0.003779
$\hat{S}_3^{25}$	0.019473	$\hat{S}_3^{57}$	0.006422
$\hat{S}_3^{26}$	0.010172	$\hat{S}_3^{58}$	0.001771
$\hat{S}_3^{27}$	0.021567	$\hat{S}_3^{59}$	0.003516
$\hat{S}_3^{28}$	0.035565	$\hat{S}_3^{60}$	0.001170
$\hat{S}_3^{29}$	0.015248	$\hat{S}_3^{61}$	0.004995
$\hat{S}_3^{30}$	0.037828	$\hat{S}_3^{62}$	0.006277
$\hat{S}_3^{31}$	0.016779	$\hat{S}_3^{63}$	0.002747

Table 4.16: Sensitivity factors for the third decomposition.

After thresholding and selecting the best basis have been done, the remaining non-zero coefficients are quantized according to each subspace, and again the quantization function is computed in the same manner as the quantization in the conventional approach. The transform coefficients are quantized by

$$\overline{X}(n_1,n_2) = \begin{cases} \left\lfloor \frac{X(n_1,n_2)-T_j^l}{Q_j^l} + 0.50 \right\rfloor & X(n_1,n_2) > 0 \\ \\ \left\lceil \frac{X(n_1,n_2)+T_j^l}{Q_j^l} - 0.50 \right\rceil & X(n_1,n_2) < 0 \end{cases} \tag{4.25}$$

where Q_j^l is calculated by

$$Q_j^l = \sqrt{A_{F_j}^l T_j^l} \tag{4.26}$$

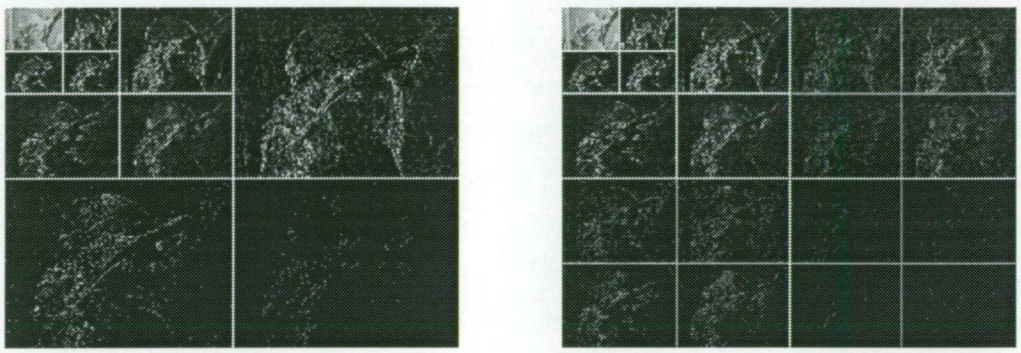
and $A_{F_j}^l$ is the activity function of the l^{th} subspace in the subband at scale 2^j.

4.4.2 Simulation Results

As stated in the previous section the simulations are only carried out for the biorthonormal spline wavelet bases with $L = 4$ and $\tilde{L} = 2$. As in Section 4.3.5 the simulation is computed for a constant bit rate of 0.40 bpp. The simulation results for both the conventional and best basis approaches with 3 levels of decomposition are tabulated in Table 4.17. Figures 4.20 and 4.21 show the selected bases in the decomposition and the reconstructed test image "Lenna" for both the conventional and best basis approaches respectively. It can be seen from Figure 4.20 that the top left quad-tree of decomposition in the best basis approach is similar to the top left quad-tree in the conventional approach. However, the top right, bottom left and bottom right quad tree decompositions in the best basis approach are different. Table 4.17 shows that the best basis approach results in a slightly higher WPSNR compared to the conventional approach. This increase is due to selecting the most appropriate basis from each decomposition level by using a minimum information cost, while in the conventional approach the choice of basis is fixed. The price we have to pay for this marginal improvement in the order of 0.6 dB for the "Lenna" image is the much more demanding computational cost. The conventional approach requires 10 operations ($3j + 1$ operations), while the best basis approach needs 64 operations (4^j operations) for three levels ($j = 3$) decomposition.

Type of approach	WPSNR [dB]
conventional	40.196
best basis	40.830

Table 4.17: Comparison of the WPSNR for the conventional and best basis approaches.



(a) (b)

Figure 4.20: The selected bases for the decomposition of the test image "Lenna", (a) for the conventional approach and (b) for the best basis approach with three levels of decomposition.

(a) (b)

Figure 4.21: The reconstruction test image "Lenna". (a) For the conventional approach and (b) for the best basis approach.

4.5 Summary

Statistical and psychovisual image compression using the wavelet transform has been investigated. The simulation results showed that the biorthonormal wavelet bases, i.e, the B-spline wavelet bases for $L = 4$ and $\tilde{L} = 2$ resulted in the highest WPSNR for both statistical and psychovisual wavelet transform coding using the conventional approach. This result was based on a mathematical measure, the WPSNR, as well as a subjective assessment performed by human observers.

The use of increasingly smoother orthonormal wavelet bases for psychovisual image compression in the conventional approach resulted in marginally higher WPSNRs. Smooth or regular wavelets do, however, need longer filters for their implementation and thus increase the computational cost of any coding scheme that utilities such wavelets. In conjunction with the HVS system, less regular wavelets result in higher sensitivity factors in the lowest subbands and lesser sensitivity factors in the higher subbands. These properties are desirable in subjective thresholding since the higher frequency ripples caused by less regular wavelets can be effectively canceled out, moreover the transform coefficients in the higher subbands can be discarded or coarsely quantized. This is supported by the fact that the increase in WPSNRs as regularity increased was only slight for orthonormal wavelets. In psychovisual image compression, the subjective thresholding and quantization are only applied on the analysis side, and therefore only the wavelet bases in the analysis side are taken into account in the HVS system. Consequently, when using biorthonormal wavelets where the analysis and synthesis filters are of different length, the less regular wavelet or short filter should be used on the analysis side. This was verified by swapping the analysis and synthesis filters and is demonstrated by the results of Table 4.10.

The conventional approach resulted in a slightly lower WPSNR compared to the best basis approach. This observation arises from the fact that the wavelet bases are fixed in the conventional approach. On the other hand, in the best basis approach, it is possible to choose the most appropriate basis for representing the given decomposed image using a minimum cost function. In terms of computational cost, however, the best basis approach is much more expensive. The comparative computational cost of the conventional and the best basis approaches is 10 and 64 operations respectively, for 3 levels decomposition. In terms of a trade-off between computational cost and WPSNR

it is clear that the conventional approach is the more optimal of the two image coding techniques.

Chapter 5

Motion Compensation Using Subjective Coded WT Coefficients

5.1 Introduction

In a number of interframe coding applications for transmitting video sequences, for example teleconferencing, the typical image sequence usually consists of a moving foreground object, usually a person, presented against a stationary background. The simplest way of viewing interframe coding is to view it as transmitting only the portion of the image which has changed, whilst maintaining the background as before.

The main difficulty in interframe coding is caused by the presence of moving objects, since this produces large differences between successive frames. The elimination of these large errors by predicting the movement or displacement of objects is known as Motion Compensation (MC). Motion compensated interframe prediction has been used as an effective technique for temporal prediction in many video coding schemes [85] [86] [87] [123] [124] [125]. In order to obtain a high rate of data compression, a combination of motion compensated interframe prediction and DCT encoding has already been implemented and standardized [6] [35] [93] [126]. All of these methods employ the Block

Matching Algorithm (BMA) for their motion compensation schemes. This chapter provides a description of basic motion compensation using a fixed block size and variable block size matching algorithms and describes and implements a video coding scheme that uses a combination of motion compensation, psychovisual compression and the wavelet transform.

5.2 Motion Compensation using Block Matching Algorithm (BMA-MC)

Motion compensation in a sequence of images has received considerable attention [82] [88] [127] [128] [129] [130]. The concept of motion compensation is to compensate motion information in a sequence of images using motion vectors. Motion compensation can be performed by "pel recursive algorithms" and "block matching algorithms". Pel recursive algorithms try to find an estimate for every pixel that has changed along the scanned line and allows the motion vectors to be determined at the receiver as well as at the transmitter, thereby bypassing the requirement to transmit the motion vectors [7, pp. 135]. The block matching algorithm requires a representative motion vector to be transmitted for every block of pixels. Since motion compensation using block matching has been implemented and standardized and is by far the more popular of the two, this study concentrates on the use of block matching rather than pel recursive algorithms.

5.2.1 Displacement Vector Detection

The principle of motion compensation employing a block matching algorithm is illustrated in Figure 5.1. The basic operation of this scheme is as follows. The current image frame is divided into fixed $N \times N$ square subblocks of pixels. A square sub-block is chosen in this study because it was found to be preferable to a rectangular block for computational reasons [131]. The BMA-MC assumes that every pixel in a subblock is a part of the same object and that the motion of this subblock is constant. Thus only two-dimensional translation is fully compensated for. Rotation, occlusion, scaling and perspective distortion cannot be compensated for completely [7, pp. 133-134]. A search is then performed in the area around each of these subblocks in the previous

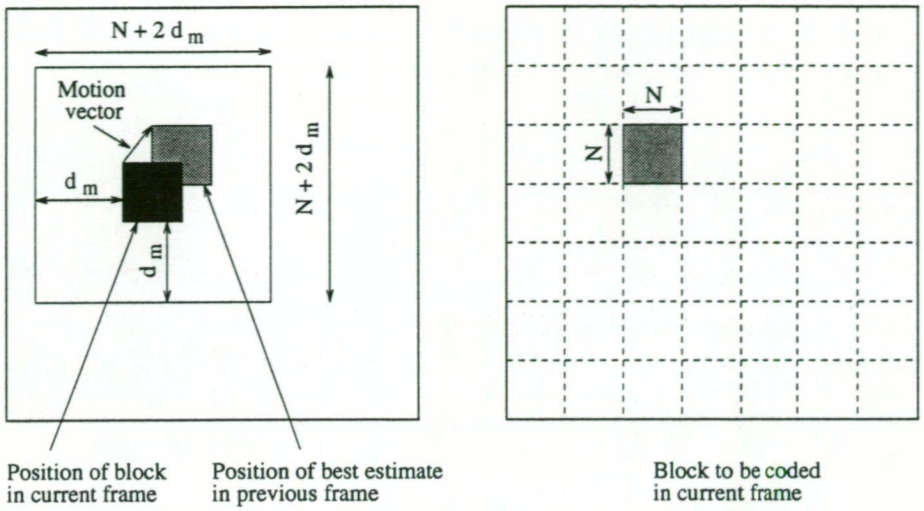


Figure 5.1: The basic Block Matching Algorithm (BMA).

frame to locate the $N \times N$ subbblock of pixels which best estimates the current subblock. This search is performed only within a limited area, i.e. less than the maximum displacement d_m pixels in the horizontal and vertical directions from the last frame.

The motion (displacement) vector of the subblock with the best fit in the previous frame then yields a measure of the displacement of that subblock which can be used to predict the corresponding subblock in the current frame. Repeating this process for all subblocks in the current image results in the construction of a block matched image, which is made up of shifted subblocks from the previous frame and represents the best estimate of the current frame from the previous frame. Transmitting the displacement vectors of these best fit subblocks and using the block matched image as a prediction of the current image frame reduces the entropy of the difference between frames to be coded and enables the overall bit rate required to encode an image sequence to be efficiently computed. Since a displacement vector is coded only once every subblock of pixels, the displacement vector then occupies a considerably smaller part of the number of bits than the number of bits representing the difference between frames or prediction error. However, redundancy of even one bit per block on average results in the number of total bits increasing by tens of Kbits per second (Kbps) [132]. There are two approaches to coding the displacement vectors, i.e., one-dimensional methods or two-dimensional methods [132]. Due to the increased number of codewords needed in the two-dimensional methods and from the viewpoint of hardware complexity, one-dimensional methods are more common for coding displacement vectors and they

usually employ a variable length coding scheme [7, pp. 152-162] [132]. One-dimensional codes for the displacement vectors are tabulated in Appendix F.

To find the best displacement vector, various measurement such as Cross Correlation Function (CCF), Mean Square Error (MSE) and Mean Absolute Difference (MAD) can be used [85] [86] [125]. These criteria are shown as follows :

$$\mathrm{CCF(x,y)} = \frac{\sum_{m=1}^{N}\sum_{n=1}^{N} \bar{f}_c(m,n)\bar{f}_p(m+x,n+y)}{(\sum_{m=1}^{N}\sum_{n=1}^{N} \bar{f}_c^2(m,n))^{\frac{1}{2}}(\sum_{m=1}^{N}\sum_{n=1}^{N} \bar{f}_p^2(m+x,n+y))^{\frac{1}{2}}} \tag{5.1}$$

$$\bar{f}_c(m,n) = f_c(m,n) - \tfrac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{N} f_c(m,n)$$
$$\bar{f}_p(m,n) = f_p(m,n) - \tfrac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{N} f_p(m,n)$$

$$\mathrm{MAD(x,y)} = \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N} |f_c(m,n) - f_p(m+x,n+y)| \tag{5.2}$$

$$\mathrm{MSE(x,y)} = \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N} (f_c(m,n) - f_p(m+x,n+y))^2 \tag{5.3}$$

where $f_c(m,n)$ represents the current subblock of $N \times N$ pixels at coordinate (m,n) of the current frame and $f_p(m+x,n+y)$ represents the corresponding subblock in the previous frame at the new coordinate $(m+x,n+y)$. It is clear that the MAD and MSE are the more convenient measures because they contain no multiplications. Because there is no significant differences in performance between MAD and MSE [89], the MAD was chosen for the purposes of this work.

5.2.2 Performance of BMA-MC

The performance of the BMA depends on several factors [7, pp. 135-151] [125]. The two most significant factors are the maximum allowable displacement and the chosen block size. A large maximum displacement is required in order to cope with large displacements, i.e., fast moving objects. Using a large maximum displacement, however, implies a longer time to find the best match. For example, if the maximum displacement is d_m pixels in the horizontal and vertical directions, the location of the best match using a full search requires the evaluation of $(2d_m + 1)^2$ matching criteria. In order to reduce the computations, various fast search methods have been proposed [85] [86] [87] [88] [89] [128]. The fast search methods, however, all result in a degradation in the reconstructed image and therefore the full search method is used in this study.

Another factor affecting the performance of the BMA scheme is the chosen block size. The optimal block size to be applied in a MC scheme depends largely on the characteristics of the image sequences being encoded. Both large and small block sizes have their advantages and disadvantages under certain conditions.

A large block size has the advantage that fewer blocks are required to cover an entire image frame, and so fewer displacement vectors need to be transmitted. However, a large block is only efficient where the motion in the frame is either non-existent or uniform in both magnitude and direction over large areas of the frame. In these cases, where the displacement is uniform over entire blocks a good estimate of the block displacement is easy to determine. This leads to less error between actual and predicted blocks and, hence, the overall data rate is reduced. If, however, the displacement is non-uniform over areas in an image such as near the boundaries of displaced objects then a large block size is a disadvantage since the block may cover different objects moving at different velocities. In these cases, it is difficult to find a good match. The block error thus rises and a higher data rate is needed for transmission.

In contrast, using a small block size means that in areas of non-uniform motion there is better matching produced. Therefore the error between blocks is lower. However, because many more blocks are required to code each frame, the contribution of the displacement vectors to the overall bit rate increases it dramatically.

5.2.3 Simulation Results

The block diagram of the simulation undertaken is illustrated in Figure 5.2. First of all the current frame and the reconstruction of the previous frame are motion compensated using the BMA-MC, to predict the matched frame from the current frame. The difference frame can be obtained by subtracting the current frame and the matched frame. The difference frame is quantised and coded. The coded difference frame is then added back to the matched frame to perform the reconstruction of the current frame. The reconstructed current frame is used as the next previous frame.

The simulation is carried out in order to evaluate the performance of the BMA-MC scheme. This performance is expressed in terms of the PSNR and the simulation is carried out using the following parameters,

- block sizes N : 2, 4, 8, 16, and 32
- maximum displacement d_m : ± 8, ± 16 and ± 32 pixels horizontally and vertically
- frame rate : 25 Hz
- five image sequences "Miss America", "Salesman", "Skiing", "Band" and "Scenic view" which have image dimensions of size 256×256

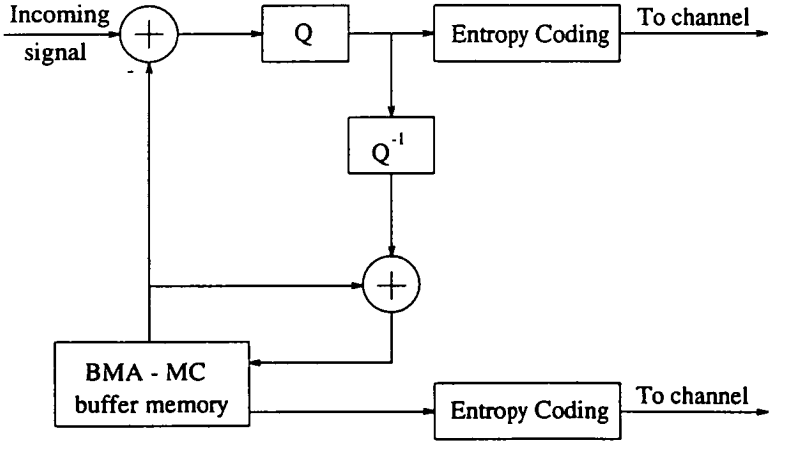


Figure 5.2: The block diagram of the simulation.

Tables 5.1, 5.2, 5.3, 5.4 and 5.5 show the simulation results, from which the following observations can be made:

1. As the block size decreases, the PSNR increases. This increase is significant for a small block size, because it is clear that a small block size must produce a more accurate match than a large block size.

2. As the maximum displacement d_m increases, the PSNR increases. Large search windows, however, result in a longer time to find the best match. This increase in PSNR is significant for large maximum displacements especially for small block sizes, because better displacement vectors can be found when using a large search window. A large search window, however, results in large displacement vectors and consequently a large number of bits for displacement vector.

3. The variation of PSNR depends on the activities of pixels moving in an image. For an image that contains uniform motion or pixels moving in the same directions, e.g., for the "Salesman" and "Skiing" images, the results of which are tabulated in Tables 5.1 and 5.2, the increase in maximum displacement gives a significant increase in the PSNR. This characteristic can be attributed to the uniform motion present within the image. If an image, however, contains much activity or has pixels moving in different directions as in the images "Miss America" and "Band", an increase in the maximum displacement does not give a significant increase in the PSNR. These characteristics are illustrated in Figure 5.3. The tables clearly show that the PSNR deteriorates steadily with increasing block size.

4. The total bits for entropy coding (total information bits) comprise the bits required for coding the displacement vectors and the bits of the difference image. A large block size results in a small number of blocks which then results in a small number of bits to represent the displacement vectors. A large block size, however, results in a large number of bits for the difference image. This is due to inaccurate predictions caused by matching large block sizes.

In summary, it is widely understood in the BMA-MC that the window size and the block size should be chosen carefully. It is clear from the simulation results that a larger window size gives better results in terms of the PSNR. From the simulation results, however, the increase in PSNR is insignificant for window sizes more than ± 8 pixels for block sizes more than 8×8. Larger windows simply result in a longer time to find the best match. For example in the test image sequence "Miss America", the use of a window size of 8 takes 42.648 seconds to code the image, a window size of 16 takes 126.412 seconds and a window size of 32 takes 454.182 seconds on a Sun SPARC

block size	d_m [pixels]	Total bit [Kbits] motion vectors	Total bit [Kbits] all information	PSNR [dB]
2 × 2	8	97.674	113.918	38.450
2 × 2	16	74.772	142.671	38.788
2 × 2	32	47.057	117.304	38.965
4 × 4	8	23.494	38.354	36.866
4 × 4	16	30.326	42.904	37.347
4 × 4	32	27.898	38.871	37.767
8 × 8	8	5.080	32.787	35.950
8 × 8	16	6.365	31.272	36.308
8 × 8	32	6.099	28.538	36.642
16 × 16	8	1.095	37.719	35.494
16 × 16	16	1.356	35.604	35.729
16 × 16	32	1.162	33.283	35.953
32 × 32	8	0.228	41.909	35.308
32 × 32	16	0.277	41.220	35.462
32 × 32	32	0.258	39.523	35.468

Table 5.1: Performance of the fixed block size BMA-MC for the image sequence "Miss America".

block size	d_m [pixels]	Total bit [Kbits] motion vectors	Total bit [Kbits] all information	PSNR [dB]
2 × 2	8	103.437	96.015	37.584
2 × 2	16	83.016	124.333	38.005
2 × 2	32	54.217	126.038	38.327
4 × 4	8	15.504	40.454	35.822
4 × 4	16	20.009	43.580	36.044
4 × 4	32	21.856	44.357	36.236
8 × 8	8	2.544	36.769	35.003
8 × 8	16	2.976	36.874	35.090
8 × 8	32	3.172	36.719	35.173
16 × 16	8	0.431	38.032	34.657
16 × 16	16	0.462	37.980	34.679
16 × 16	32	0.470	37.903	34.716
32 × 32	8	0.072	41.106	34.552
32 × 32	16	0.072	41.106	34.552
32 × 32	32	0.072	41.106	34.552

Table 5.2: Performance of the fixed block size BMA-MC for the image sequence "Salesman".

block size	d_m [pixels]	Total bit [Kbits] motion vectors	Total bit [Kbits] all information	PSNR [dB]
2×2	8	101.969	115.214	37.802
2×2	16	81.078	138.495	38.293
2×2	32	52.821	126.275	38.584
4×4	8	19.642	60.797	36.441
4×4	16	25.162	64.093	36.798
4×4	32	24.830	61.809	37.081
8×8	8	3.762	59.849	35.738
8×8	16	4.507	59.600	35.971
8×8	32	4.373	58.791	36.165
16×16	8	0.685	65.494	35.406
16×16	16	0.784	64.786	35.513
16×16	32	0.732	64.298	35.664
32×32	8	0.123	68.825	35.199
32×32	16	0.139	68.970	35.254
32×32	32	0.141	69.127	35.336

Table 5.3: Performance of the fixed block size BMA-MC for the image sequence "Band".

block size	d_m [pixels]	Total bit [Kbits] motion vectors	Total bit [Kbits] all information	PSNR [dB]
2×2	8	103.214	104.482	37.862
2×2	16	83.266	129.423	38.274
2×2	32	55.681	125.804	38.589
4×4	8	17.573	53.590	36.491
4×4	16	22.133	56.056	36.669
4×4	32	22.269	54.834	36.825
8×8	8	3.273	51.271	35.673
8×8	16	3.965	50.685	35.791
8×8	32	3.777	49.566	35.899
16×16	8	0.584	53.918	35.330
16×16	16	0.676	53.628	35.387
16×16	32	0.636	53.327	35.417
32×32	8	0.083	56.547	35.192
32×32	16	0.086	56.559	35.193
32×32	32	0.084	56.389	35.206

Table 5.4: Performance of the fixed block size BMA-MC for the image sequence "Skiing".

block size	d_m [pixels]	Total bit [Kbits] motion vectors	Total bit [Kbits] all information	PSNR [dB]
2×2	8	106.790	116.830	37.347
2×2	16	85.249	146.466	37.775
2×2	32	54.562	124.647	38.122
4×4	8	21.842	74.870	35.766
4×4	16	28.619	77.934	35.852
4×4	32	26.549	75.284	35.963
8×8	8	4.280	72.074	35.284
8×8	16	5.086	72.389	35.305
8×8	32	4.857	72.176	35.334
16×16	8	0.956	73.732	35.124
16×16	16	1.085	73.679	35.143
16×16	32	0.970	73.387	35.158
32×32	8	0.225	76.227	35.053
32×32	16	0.249	76.212	35.067
32×32	32	0.204	76.101	35.076

Table 5.5: Performance of the fixed block size BMA-MC for the image sequence "Scenic view".

station IPX operating under conditions of no load. These results clearly indicate that a window size of ± 8 pixels in both the horizontal and vertical directions is a good trade-off between PSNR and computational expense. This supports the previous research performed in [7, pp. 133-162].

Small block sizes result in a higher PSNR, since a more accurate match can be obtained. Small block sizes require, however, more displacement vectors for the larger number of blocks and in this case results in a higher number of total bits. The tables and above mentioned computational times show that when using a block size 8×8 with a window size of 8 pixels in both the horizontal and vertical directions an optimum total information bit rate is obtained. When using the same window size of 8 pixels in both the horizontal and vertical directions, the PSNR is less compared to block sizes of 4×4 and 2×2 reinforcing the necessity for a trade-off between block size and computational time when choosing a suitable block size. Large block sizes are advantageous in areas of uniform motion, small block sizes are necessary in active areas or where non-uniform motion is present. These trends also suggest that a variable block size could be an advantage in a block matching motion compensation scheme.

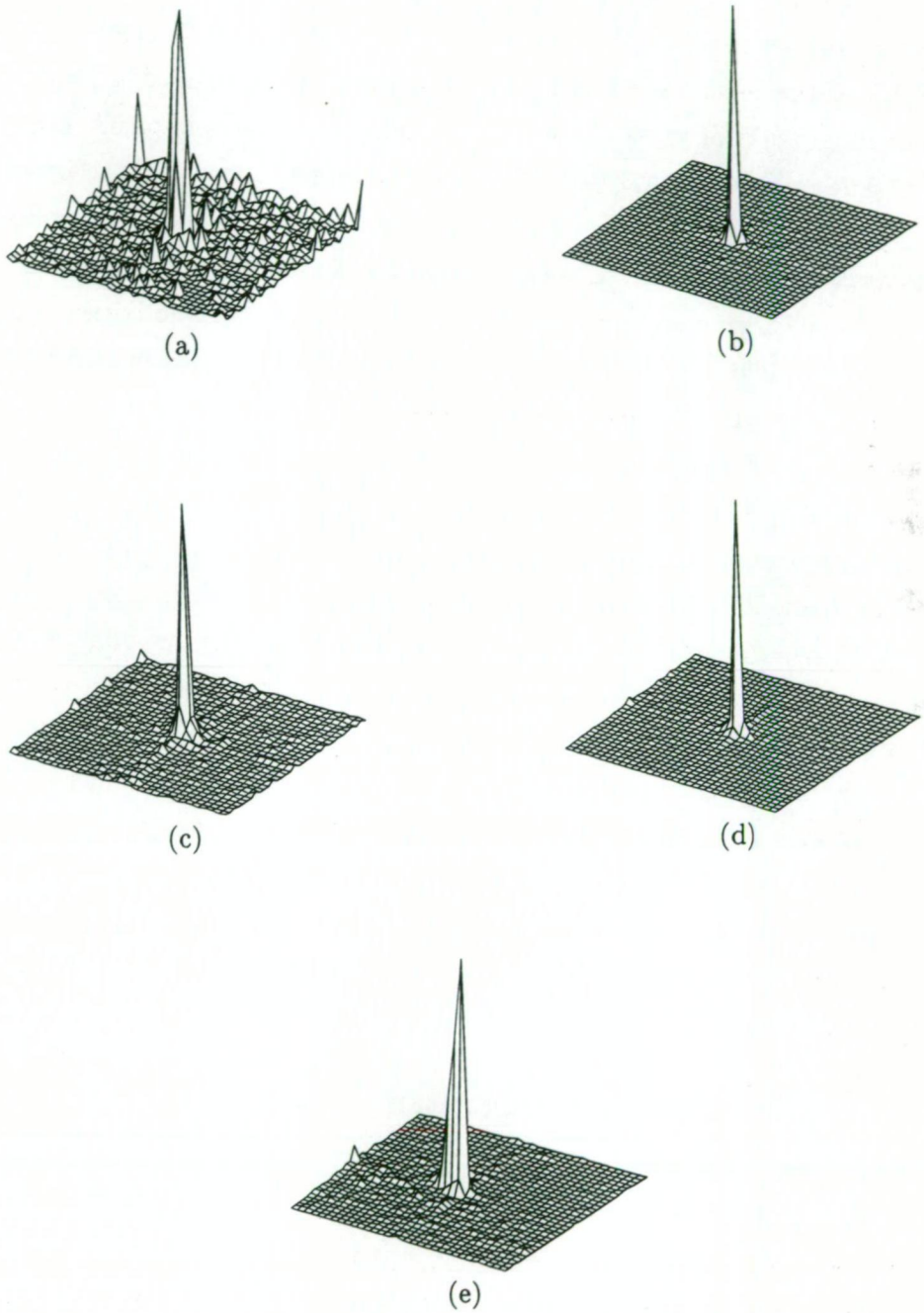


Figure 5.3: Characteristic displacement vectors for the test image (a) "Miss America", (b) "Salesman", (c) "Band", (d) "Skiing" and (e) "Scenic view".

5.3 The variable size BMA-MC

The variable size block BMA-MC scheme is similar to the conventional fixed block BMA in that a best match is sought between the previous and current frames [4]. In the fixed block BMA, the displacement vector is found using the minimum MAD value obtained by matching blocks of pixels in the current frame with blocks of pixels in the previous frame. The minimum MAD value is zero if the current block and the previous block are perfectly matched or if there is no difference (no translation or moving pixels in the block) between the previous and the current blocks. In general image sequences, the MAD value is larger than zero, especially in blocks with high activities. Appendix G illustrates the minimum MAD values for the test image sequences for varying block sizes and a maximum window size of ± 8 pixels in both the horizontal and vertical directions. As with the fixed BMA these MAD values are used to determine the displacement vectors. In Appendix G a large MAD value is due to poor matching resulting in less accurate displacement vectors. In the variable size block matching scheme, a given threshold value is employed. This value is used for thresholding the MAD values of each block when a search is being performed. If the MAD value of a certain block is larger than the given threshold value, this block size should be reduced by a factor of two, dividing the original block into four blocks which have a block size half the original block. Matching is again performed with these four blocks to find the best block match. This process of subdivision continues until a MAD value less than the given threshold is achieved or a given minimum block size is reached. As can be seen the variable block size matching scheme is similar to the conventional fixed block matching scheme, except that size of the blocks is now dependent on the characteristics of the image.

5.3.1 The principle of the algorithm

The basic operation of the variable size BMA-MC scheme is as follows [4]. Initially, the largest allowed square block of size $N = N_{max}$, $N_{max} \times N_{max}$ of pixels is employed to encode an image. This large block size is used for predicting the stationary areas or objects such as the background. In the block matching, the minimum value of MAD is computed and compared to the given threshold value. If the MAD value is less than or equal to the given threshold value then a match is obtained. If a suitable best fit was found for this block in the previous frame, then this estimate is used for the

displacement vector and the next block in the image is processed. If, however, the minimum MAD value is larger than the given threshold value, then this implies that there is too much activity within the block and a smaller block size needs to be used. The block size is then reduced by a factor of four to $\frac{N}{2} \times \frac{N}{2}$. Four blocks of this size are then required to encode the original $N \times N$ block. Matches for each of these blocks are then performed by searching in the previous frame. This process is repeated until either an adequate match is found for each block or until the block size is reduced to the smallest allowable size $N_{min} \times N_{min}$. If the minimum MAD of the $N_{min} \times N_{min}$ sized blocks are still greater than the given threshold value, then there is no alternative but to use this size block, with a less than acceptable match.

Figure 5.4 illustrates a possible resulting structure of the block matched image, depicting a typical "patch work" characteristic. Areas of uniform motion such as sections of non-moving background are coded using larger blocks, but as the uniformity of the motion decreases, such as around the edges of a moving object, the block size is reduced accordingly.

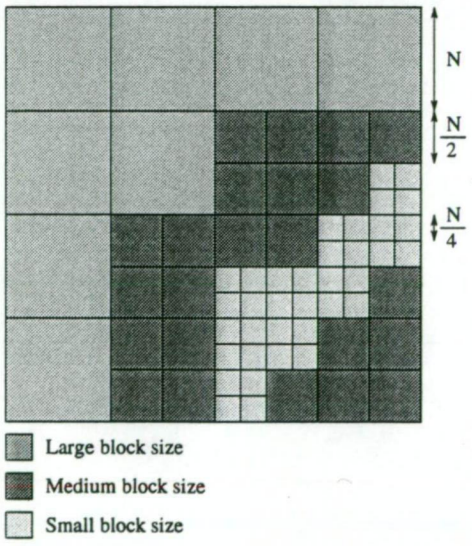


Figure 5.4: The "patch-work" characteristic of the variable block matching algorithm.

5.3.2 Coding Requirements

In the variable block BMA-MC technique, a large block size can be further divided into small block sizes when the MAD value is larger than a given threshold. To distinguish between the subblock $N \times N$ and the divided subblock $\frac{N}{2} \times \frac{N}{2}$, it is necessary to allocate one control bit to represent the block size. This can be realised by using the flag "1" for a subblock which has been subdivided and "0" for subblocks that require no further subdivision. Take for example, a frame size of $M \times M$ pixels, a maximum block size of $N_{max} \times N_{max}$, a minimum block size of $N_{min} \times N_{min}$. The best situation occurs when there is no subdivision of $N_{max} \times N_{max}$. This means the total number of control bits is $(\frac{M}{N_{max}})^2$. In the worse case, where all the blocks are divided into $N_{min} \times N_{min}$ sized blocks, the total number of control bits is $(\frac{2M}{N_{max}})^2$ if the minimum block size is $N_{min} = \frac{N_{max}}{2}$. Therefore, the total range for the number of control bits is from $(\frac{2M}{N_{max}})^2$ to $(\frac{M}{N_{max}})^2$ bits. The total number of transmitted bits for the whole frame can then be obtained by summing the entropy coding of the difference frame, the entropy coding of the displacement vectors and the total number of control bits. Figure 5.5 shows a possible subdivision of the frame.

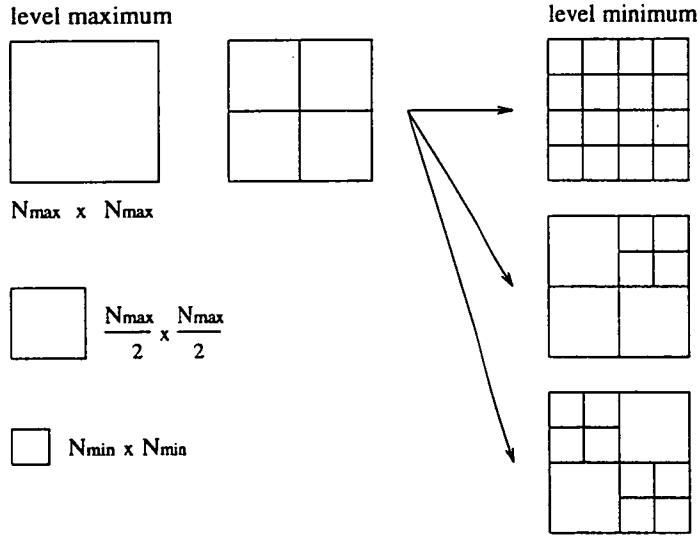


Figure 5.5: A possible subdivision of an original frame.

5.3.3 Simulation Results

The simulation is carried out using a combination of variable block sizes ranging from a maximum block size of 32 × 32 to a minimum block size of 2 × 2. Tables 5.6, 5.7, 5.8, 5.9 and 5.10 show the simulation results for the image sequences "Miss America", "Salesman", "Skiing", "Band" and "Scenic view" respectively. Figure 5.6 shows example of the "Miss America" image subdivided into blocks varying between a minimum size of 2 × 2 and a maximum size of 8 × 8 for thresholds of 3 to 7. Figures 5.7 and 5.8 also show the results for the test images "Salesman", "Band", "Skiing" and "Scenic view" for threshold value equal to 3. From Figure 5.6, it can be seen that a subblock which has a MAD value larger than the given threshold results in greater subdivision of the block. For a threshold of 3, the subblocks are, in general, more subdivided than for a threshold of 6 or 7.

From the simulation results the following observations can be made:

1. All the tables clearly show that an increase in the threshold value decreases the number of total bits for thresholds less than or equal to 5, and for thresholds larger than 5 results in a slight increase especially when using a maximum block size more than 8. The higher threshold value results in a lesser number of total control bits and displacement vectors, but more bits are needed for transmitting the difference image (prediction error) since the higher threshold values result in poorer matching. By increasing the threshold value, the total number of displacement vectors and control bits is decreased, as is the PSNR.

2. The combination of maximum size 8 × 8, minimum size 2 × 2 and a threshold value of 6 resulted in the minimum number of totals bits for the images "Miss America", "Salesman" and "Scenic view". The threshold value of 7 resulted in the minimum number of totals bits for the images "Band" and "Skiing". This difference between threshold values arises from the different MAD values which are themselves a result of the different levels of activity between the images, as illustrated in Appendix G. If an image has a lot of low minimum MAD values in its subblocks, a higher threshold value results in a more significant reduction in the number of total bits, as demonstrated for the test images "Salesman" and "Skiing". A low minimum MAD value in the subblock image represents a good match in the search. If an image, however, has a lot of high minimum

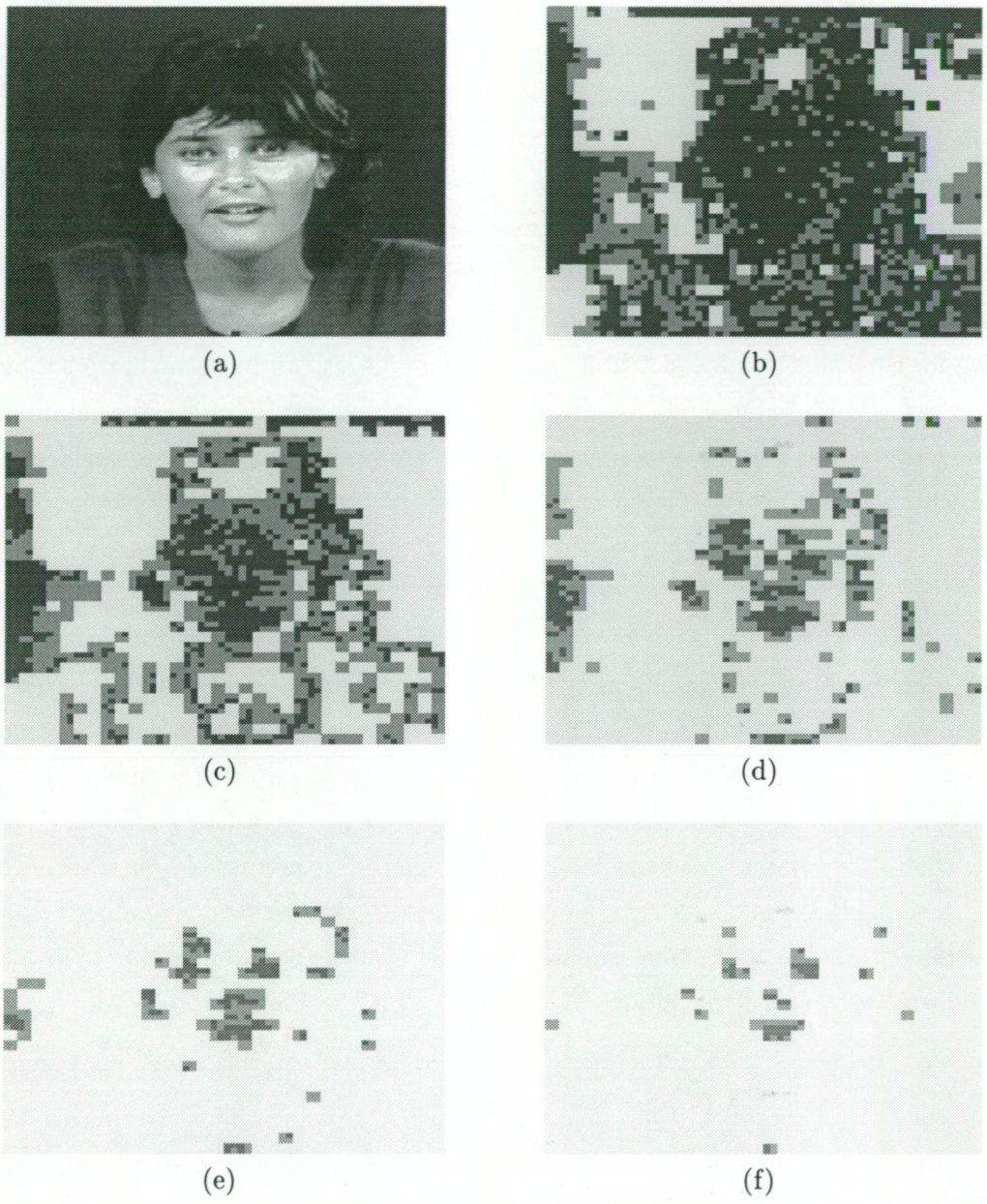
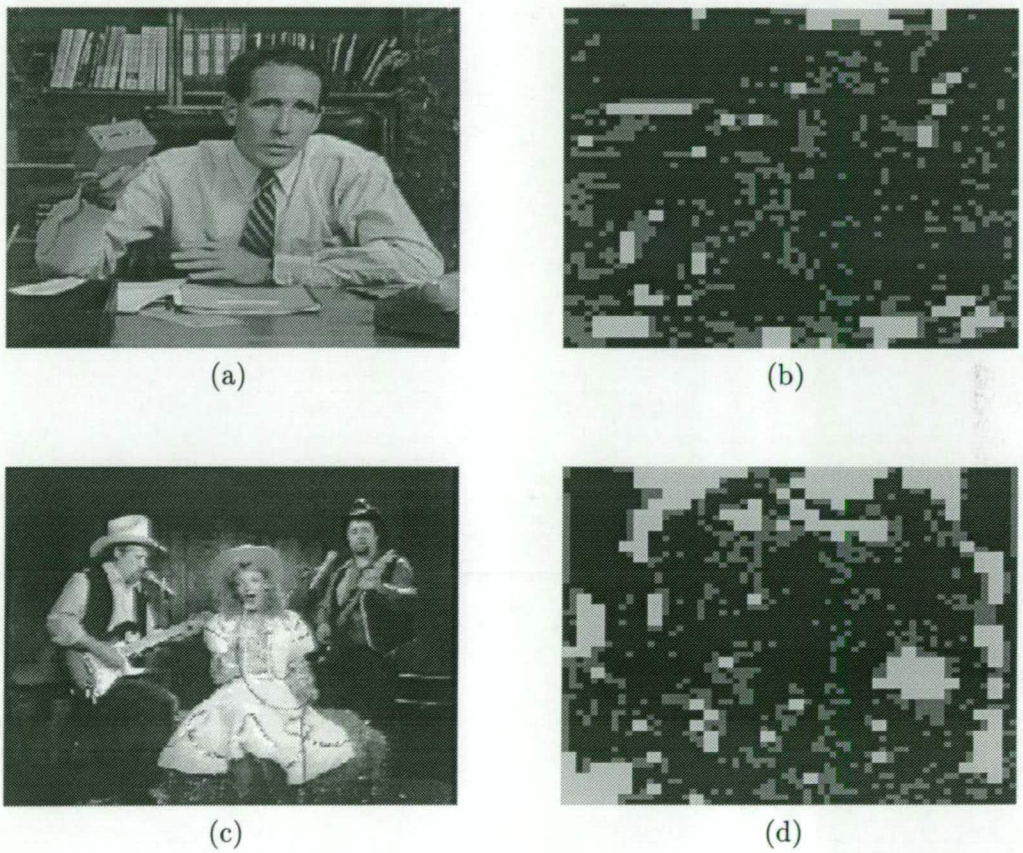


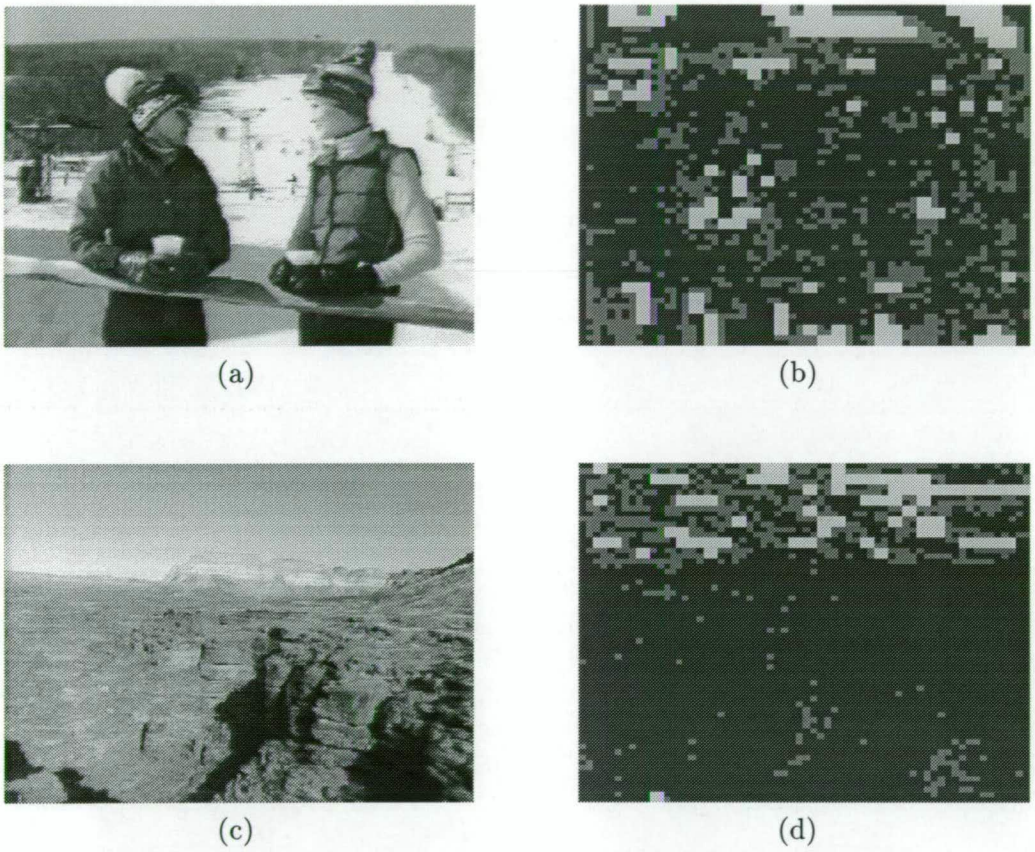
Figure 5.6: Examples of the subdivision of the test image sequence "Miss America" into a maximum block size of 8×8 *and a minimum block size of* 2×2 *in the variable block size block matching algorithm. (a) The original image, (b) an example for a threshold value = 3, (c) an example for a threshold = 4, (d) an example for a threshold value = 5, (e) an example for a threshold value = 6 and (f) an example for a threshold value = 7. The brightest colour represents a block size of* 8×8*, the next brightest represents a block size of* 4×4 *and the darkest colour represents a block size of* 2×2*.*



(a) (b)

(c) (d)

Figure 5.7: Examples of subdivision for the test image sequences "Salesman" and "Band' into a maximum block size of 8×8 *and a minimum block size of* 2×2 *in the variable block size matching algorithm. (a) The original image sequence "Salesman", (b) an example for a threshold value = 3, (c) the original image sequence "Band" and (d) an example for a threshold value = 3. The brightest colour represents a block size of* 8×8*, the next brightest represents a block size of* 4×4 *and the darkest colour represents a block size of* 2×2*.*



(a) (b)

(c) (d)

Figure 5.8: Examples of subdivision for the test image sequences "Skiing" and "Scenic view" into a maximum block size of 8×8 *and a minimum block size of* 2×2 *in the variable block size matching algorithm. (a) The original image sequence "Skiing", (b) an example for a threshold value = 3, (c) the original image sequence "Scenic view" and (d) an example for a threshold value = 3. The brightest colour represents a block size of* 8×8*, the next brightest represents a block size of* 4×4 *and the darkest colour represents a block size of* 2×2*.*

maximum block	minimum block	Threshold value	control bits [bits]	Total bit [Kbits]	PSNR [dB]
8 × 8	2 × 2	3	12692	76.370	37.983
8 × 8	2 × 2	4	6416	44.769	37.060
8 × 8	2 × 2	5	2904	32.706	36.232
8 × 8	2 × 2	6	1696	30.580	36.046
8 × 8	2 × 2	7	1276	30.701	35.968
8 × 8	4 × 4	3	4028	35.354	36.745
8 × 8	4 × 4	4	2976	32.653	36.506
8 × 8	4 × 4	5	1780	31.100	36.165
8 × 8	4 × 4	6	1288	30.993	36.002
8 × 8	4 × 4	7	1120	31.223	35.958
16 × 16	2 × 2	3	12676	77.670	37.941
16 × 16	2 × 2	4	6024	45.691	36.903
16 × 16	2 × 2	5	2276	35.312	36.003
16 × 16	2 × 2	6	856	35.045	35.590
16 × 16	2 × 2	7	480	35.586	35.508
16 × 16	4 × 4	3	4096	36.265	36.716
16 × 16	4 × 4	4	2748	34.703	36.377
16 × 16	4 × 4	5	1260	34.142	35.871
16 × 16	4 × 4	6	532	35.510	35.559
16 × 16	4 × 4	7	360	36.161	35.502
32 × 32	2 × 2	3	12712	78.303	37.879
32 × 32	2 × 2	4	5980	47.180	36.834
32 × 32	2 × 2	5	1892	38.757	35.769
32 × 32	2 × 2	6	512	40.359	35.388
32 × 32	2 × 2	7	212	41.175	35.316
32 × 32	4 × 4	3	4148	36.969	36.682
32 × 32	4 × 4	4	2744	36.240	36.308
32 × 32	4 × 4	5	980	37.916	35.657
32 × 32	4 × 4	6	280	40.752	35.359
32 × 32	4 × 4	7	132	41.471	35.313

Table 5.6: Results for the variable block size BMA-MC for the image sequence "Miss America".

maximum block	minimum block	Threshold value	control bits [bits]	Total bit [Kbits]	PSNR [dB]
8 × 8	2 × 2	3	17376	108.524	37.380
8 × 8	2 × 2	4	12008	77.789	36.698
8 × 8	2 × 2	5	5148	47.623	35.641
8 × 8	2 × 2	6	1832	37.208	35.105
8 × 8	2 × 2	7	1332	36.721	35.030
8 × 8	4 × 4	3	4840	46.720	35.774
8 × 8	4 × 4	4	4204	44.286	35.645
8 × 8	4 × 4	5	2596	40.046	35.329
8 × 8	4 × 4	6	1352	36.937	35.063
8 × 8	4 × 4	7	1136	36.946	35.018
16 × 16	2 × 2	3	17612	108.692	37.376
16 × 16	2 × 2	4	12004	77.519	36.615
16 × 16	2 × 2	5	4068	44.382	35.250
16 × 16	2 × 2	6	912	36.264	34.724
16 × 16	2 × 2	7	508	36.615	34.675
16 × 16	4 × 4	3	5088	46.955	35.771
16 × 16	4 × 4	4	4296	44.268	35.590
16 × 16	4 × 4	5	1984	38.922	35.023
16 × 16	4 × 4	6	552	36.533	34.699
16 × 16	4 × 4	7	368	36.917	34.667
32 × 32	2 × 2	3	17676	108.918	37.348
32 × 32	2 × 2	4	12068	77.720	36.594
32 × 32	2 × 2	5	3456	43.147	35.103
32 × 32	2 × 2	6	576	39.062	34.576
32 × 32	2 × 2	7	196	40.018	34.552
32 × 32	4 × 4	3	5152	47.347	35.749
32 × 32	4 × 4	4	4360	44.684	35.563
32 × 32	4 × 4	5	1688	38.812	34.924
32 × 32	4 × 4	6	292	39.244	34.564
32 × 32	4 × 4	7	120	40.011	34.551

Table 5.7: Results for the variable block size BMA-MC for the image sequence "Salesman".

maximum block	minimum block	Threshold value	control bits [bits]	Total bit [Kbits]	PSNR [dB]
8 × 8	2 × 2	3	15028	101.697	37.544
8 × 8	2 × 2	4	10104	74.394	36.928
8 × 8	2 × 2	5	6472	60.420	36.355
8 × 8	2 × 2	6	4492	54.635	36.057
8 × 8	2 × 2	7	3476	53.709	35.938
8 × 8	4 × 4	3	4420	61.726	36.380
8 × 8	4 × 4	4	3600	58.676	36.220
8 × 8	4 × 4	5	2680	56.968	36.002
8 × 8	4 × 4	6	2124	56.295	35.875
8 × 8	4 × 4	7	1836	56.566	35.830
16 × 16	2 × 2	3	15056	101.765	37.519
16 × 16	2 × 2	4	9812	74.393	36.792
16 × 16	2 × 2	5	5928	60.564	36.101
16 × 16	2 × 2	6	3628	56.381	35.701
16 × 16	2 × 2	7	2504	56.062	35.593
16 × 16	4 × 4	3	4552	61.788	36.365
16 × 16	4 × 4	4	3508	58.838	36.126
16 × 16	4 × 4	5	2348	57.725	35.788
16 × 16	4 × 4	6	1508	58.095	35.550
16 × 16	4 × 4	7	1152	58.641	35.505
32 × 32	2 × 2	3	15004	100.995	37.442
32 × 32	2 × 2	4	9712	74.348	36.707
32 × 32	2 × 2	5	5672	61.588	35.945
32 × 32	2 × 2	6	3324	58.038	35.452
32 × 32	2 × 2	7	2008	59.285	35.334
32 × 32	4 × 4	3	4552	61.788	36.365
32 × 32	4 × 4	4	3508	58.838	36.126
32 × 32	2 × 2	5	2212	58.779	35.646
32 × 32	4 × 4	6	1312	60.048	35.313
32 × 32	4 × 4	7	852	61.284	35.261

Table 5.8: Results for the variable block size BMA-MC for the image sequence "Band".

maximum block	minimum block	Threshold value	control bits [bits]	Total bit [Kbits]	PSNR [dB]
8 × 8	2 × 2	3	15588	104.103	37.569
8 × 8	2 × 2	4	10648	77.708	36.926
8 × 8	2 × 2	5	6060	59.969	36.252
8 × 8	2 × 2	6	3280	52.496	35.888
8 × 8	2 × 2	7	2016	50.706	35.755
8 × 8	4 × 4	3	4672	57.493	36.372
8 × 8	4 × 4	4	3888	55.579	36.200
8 × 8	4 × 4	5	2696	53.058	35.949
8 × 8	4 × 4	6	1788	51.640	35.790
8 × 8	4 × 4	7	1352	51.357	35.726
16 × 16	2 × 2	3	15760	104.325	37.557
16 × 16	2 × 2	4	10684	77.235	36.893
16 × 16	2 × 2	5	5516	59.841	36.036
16 × 16	2 × 2	6	2044	52.525	35.520
16 × 16	2 × 2	7	756	51.758	35.363
16 × 16	4 × 4	3	4880	57.660	36.365
16 × 16	4 × 4	4	4008	55.717	36.178
16 × 16	4 × 4	5	2388	53.384	35.769
16 × 16	4 × 4	6	972	52.574	35.462
16 × 16	4 × 4	7	440	52.731	35.348
32 × 32	2 × 2	3	15824	104.730	37.533
32 × 32	2 × 2	4	10704	77.396	36.863
32 × 32	2 × 2	5	5328	60.172	35.928
32 × 32	2 × 2	6	904	54.422	35.247
32 × 32	2 × 2	7	396	54.625	35.209
32 × 32	4 × 4	3	4944	57.975	36.348
32 × 32	4 × 4	4	4056	56.047	36.150
32 × 32	4 × 4	5	2304	54.450	35.669
32 × 32	4 × 4	6	392	55.228	35.217
32 × 32	4 × 4	7	192	55.490	35.195

Table 5.9: Results for the variable block size BMA-MC for the image sequence "Skiing".

maximum block	minimum block	Threshold value	control bits [bits]	Total bit [Kbits]	PSNR [dB]
8 × 8	2 × 2	3	17992	122.161	37.179
8 × 8	2 × 2	4	13468	95.190	36.627
8 × 8	2 × 2	5	10316	81.799	36.217
8 × 8	2 × 2	6	7156	72.169	35.833
8 × 8	2 × 2	7	4904	68.301	35.607
8 × 8	4 × 4	3	4852	77.763	35.734
8 × 8	4 × 4	4	3992	73.811	35.564
8 × 8	4 × 4	5	3428	72.100	35.477
8 × 8	4 × 4	6	2784	70.713	35.409
8 × 8	4 × 4	7	2208	70.068	35.377
16 × 16	2 × 2	3	18208	122.285	37.171
16 × 16	2 × 2	4	13472	95.629	36.517
16 × 16	2 × 2	5	9944	81.790	36.043
16 × 16	2 × 2	6	6592	72.209	35.661
16 × 16	2 × 2	7	4036	68.626	35.444
16 × 16	4 × 4	3	5084	77.972	35.729
16 × 16	4 × 4	4	4008	73.647	35.479
16 × 16	4 × 4	5	3232	71.754	35.350
16 × 16	4 × 4	6	2420	70.546	35.270
16 × 16	4 × 4	7	1616	70.114	35.237
32 × 32	2 × 2	3	18272	122.520	37.153
32 × 32	2 × 2	4	13440	96.107	36.438
32 × 32	2 × 2	5	9856	82.780	35.961
32 × 32	2 × 2	6	6432	73.259	35.572
32 × 32	2 × 2	7	3572	69.422	35.332
32 × 32	4 × 4	3	5148	78.252	35.700
32 × 32	4 × 4	4	3988	74.150	35.417
32 × 32	4 × 4	5	3184	72.657	35.283
32 × 32	4 × 4	6	2316	71.290	35.202
32 × 32	4 × 4	7	1352	71.271	35.148

Table 5.10: Results for the variable block size BMA-MC for the image sequence "Scenic view".

MAD values in its subblocks, then a higher threshold value does not result in a significant increase in the number of total bits, as demonstrated for the test images "Scenic view" and "Band".

In summary, it is clear that the maximum block sizes of 8×8 and the minimum block sizes of 2×2 with a threshold value of 6 or 7 result in the optimum number of total bits for the variable block size BMA-MC scheme. The maximum block size of 8×8 is used so as to limit the number of bits representing the displacement vectors and the minimum block size of 2×2 is used so as to obtain an improvement in the quality of the reconstructed images compared to the fixed block size BMA. Although an improvement over the fixed block size BMA the overall coding can be significantly improved further by incorporating transform coding. As shown in the next section this also allows the use of a lower threshold value.

5.4 Motion-compensated Wavelet Transform Coding

In order to produce a higher compression ratio using motion compensation a combination of motion compensation and the transform coding scheme can be used. This technique is called interframe hybrid motion compensation. A general block diagram of the scheme is shown in Figure 5.9. A motion vector is detected by either the fixed block size or variable block size matching algorithm. Prediction error is obtained by taking the difference between the incoming frame and the displaced frame signals. The prediction error is then divided into small blocks and transformed with a 2-D forward transform and quantised. Motion vectors and quantized coefficients are code-converted into Huffman code using variable word lengths in the entropy coding part. In existing standard interframe hybrid motion compensation schemes (Figure 5.9), the prediction error is transformed using the DCT, which unfortunately results in a blocking effect for low bit rates. It can be seen from the simulation results in the previous chapter that the multiresolution decomposition wavelet transform avoided the blocking effect for low bit rate applications. If the wavelet transform is used to replace the DCT in the interframe hybrid motion compensation scheme we can effectively eliminate this undesirable effect. This section describes the interframe hybrid multiresolution motion-compensated scheme where the multiresolution wavelet decomposition is applied either before or after motion compensation.

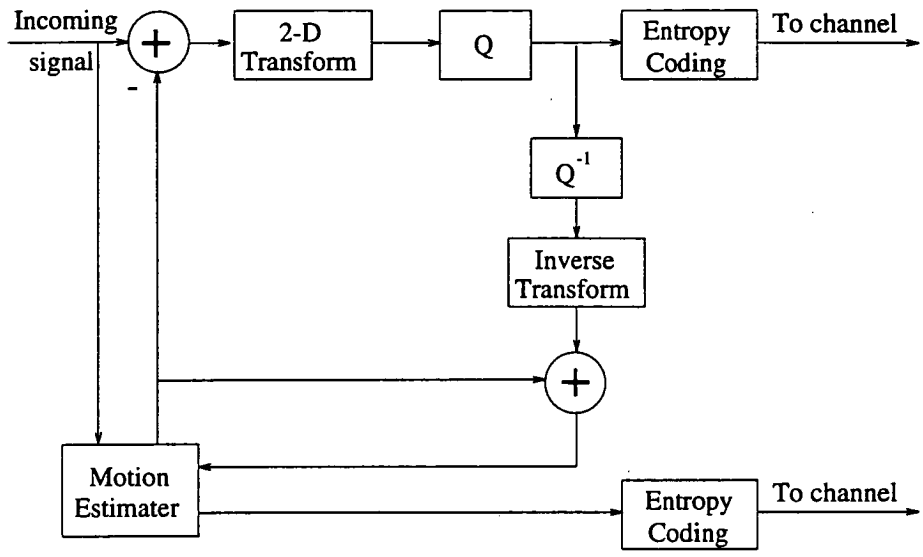


Figure 5.9: The block diagram of hybrid motion-compensated scheme.

5.4.1 Multiresolution Motion Compensation

In the interframe hybrid multiresolution motion compensation scheme, the multiresolution wavelet decomposition can be applied to either the original video samples before the motion compensation or to the residual video samples after motion compensation. The latter scheme is the same as conventional interframe hybrid motion-compensated coding except that the DCT transform [35] [93] [126] is replaced by the wavelet transform. Since the multiresolution property of the wavelet transform allows the scheme to be adapted to the HVS system, and that the variable block size matching algorithm is superior compared to the fixed block size matching algorithm, we therefore propose a variable block matching scheme where the wavelet transform coefficients are psychovisually thresholded using the HVS.

In a multiresolution image decomposition using the conventional approach, an image is decomposed into a set of quad-tree $(3j + 1)$ subbands with different resolutions as illustrated in Figure 4.3 for the case $j = 3$. These multiresolution subbands provide a representation of the motion structure at difference scales. Since the same filters are used in the multiresolution scheme, the motion activities for the subbands are highly correlated since they actually specify the same motion structure at different scales [133] [134]. Therefore, the motion vector can be firstly estimated for the lowest-resolution

subbands. Then, motion vectors in the higher-resolution subbands are predicted by using the motion vectors in the lower-resolution subbands and are refined at each step [134] [135]. This scheme results in a reduction in the searching and matching times, but the overall reconstructed image is of lesser quality compared to the use of all subbands as performed in the block matching algorithms [134]. This is because any error generated by the motion compensation in the lowest-resolution subbands will be propagated and expanded to all subsequent higher-resolution subbands. This scheme is illustrated in Figure 5.10.

The block diagram of the proposed scheme is shown in Figure 5.11. The multiresolution wavelet transform is first carried out for the incoming input image sequences. The output of each subband, the wavelet transform coefficients, are subjectively thresholded according to the HVS sensitivity function, as described in the Section 4.3.2. These outputs are then motion-compensated. The variable block size matching algorithm used is based on the method described in Section 5.3. In order to improve the quality of the reconstructed image, the full motion estimation scheme is used in this study, in which each subband of the multiresolution motion compensation is treated in similar manner to the conventional variable block size motion compensation scheme. This proposed scheme differs from the recently proposed application of motion compensation to multiresolution analysis proposed by Zhang et al. in two ways [134] [5]. Firstly, their scheme does not incorporate the human visual system in their block matching decisions. Secondly, their block size, whilst varying with the scale employed in the wavelet transform, is fixed at a particular scale. The algorithm used here determines the block size solely on how well the match is made, regardless of the scale of the transform where the block is being matched.

5.4.2 Simulation Results

The interframe hybrid multiresolution motion compensation scheme is implemented using the variable size block matching algorithm. Only the biorthonormal spline wavelet basis for $L = 4$ and $\tilde{L} = 2$ is used for the multiresolution wavelet transform. The existing interframe hybrid schemes using both the DCT and the wavelet transforms, where the multiresolution wavelet decomposition is applied to the residual video samples after motion compensation, are also simulated using the fixed block size block matching algorithm in order to compare the two schemes. Tables 5.11 and 5.12 show

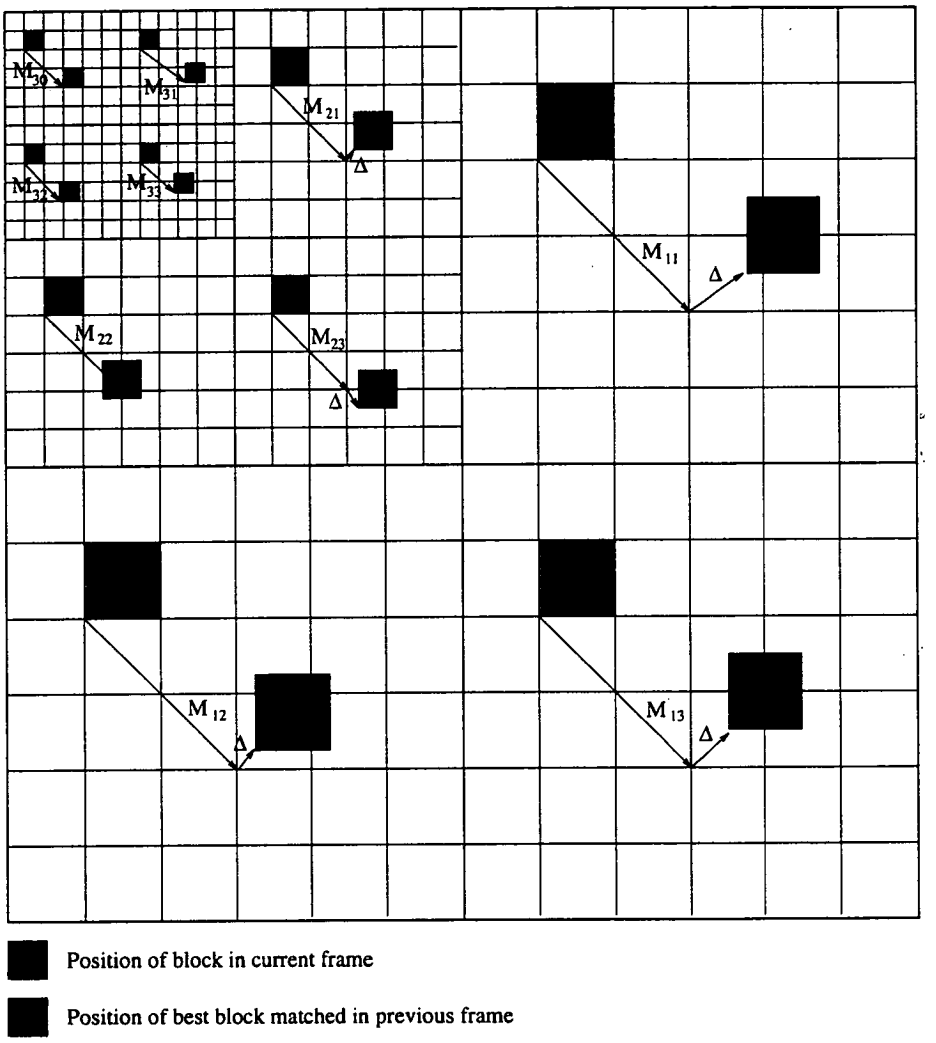


Figure 5.10: A typical multiresolution motion estimation using a scaled block size. M denotes a motion vector, the numbers 30, 31, 32 and 33 correspond to the subbands HH, GH, HG and GG at the third decomposition respectively. The numbers 21, 22 and 23 for the second decomposition and 11, 12 and 13 for the first decomposition. Δ denotes the difference in the position of the motion vector from the actual position.

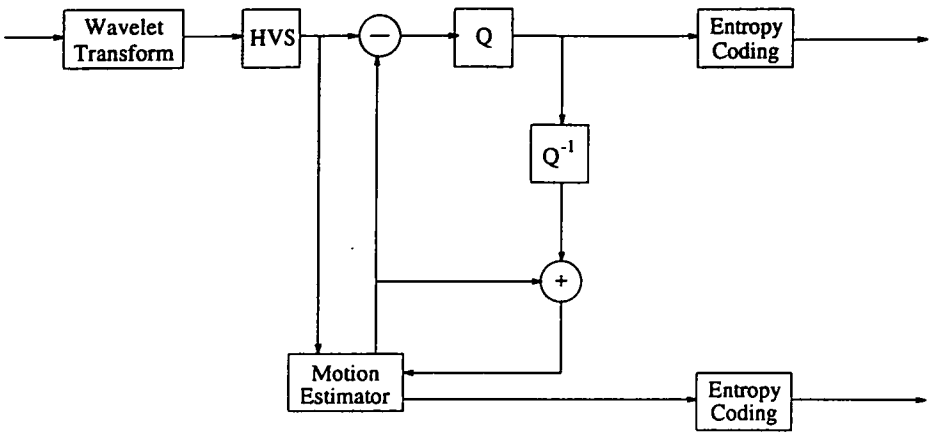


Figure 5.11: The block diagram of the proposed scheme.

the simulation results for the existing schemes using the test image sequences "Miss America", "Salesman", "Band", "Skiing" and "Scenic view". Tables 5.13, 5.14, 5.15, 5.16 and 5.17 show the simulation results for the proposed scheme, where the wavelet transform is applied to the original video samples and then motion compensated. The total bits are the number of total bits transmitted which comprise the bits required for coding the displacement vectors, the bits of control subdivided blocks and the bits of the difference image.

test image	Total bit [Kbits]	WPSNR [dB]
"Miss America"	12.045	42.661
"Salesman"	8.672	41.807
"Band"	26.080	40.932
"Skiing"	17.371	41.190
"Scenic view"	34.161	39.187

Table 5.11: Performance of the existing scheme using the DCT transform as shown in Figure 5.9.

test image	Total bit [Kbits]	WPSNR [dB]
"Miss America"	11.265	43.707
"Salesman"	7.337	42.217
"Band"	23.557	41.039
"Skiing"	15.180	41.107
"Scenic view"	29.298	39.237

Table 5.12: Performance of the existing scheme using the wavelet transform. This scheme is similar to that proposed by Zhang [133] except that a constant step size in the quantization is used so as to allow a comparison with our method that incorporates the HVS.

maximum block size	minimum block size	Threshold value	Total bit [Kbits]	WPSNR [dB]
8×8	2×2	3	9.461	44.813
8×8	2×2	4	8.491	44.216
8×8	2×2	5	7.518	43.609
8×8	2×2	6	8.091	43.353
8×8	2×2	7	7.955	43.139

Table 5.13: Performance of our proposed scheme for the image sequence "Miss America".

maximum block size	minimum block size	Threshold value	Total bit [Kbits]	WPSNR [dB]
8×8	2×2	3	6.341	43.088
8×8	2×2	4	5.910	42.772
8×8	2×2	5	5.487	42.490
8×8	2×2	6	5.852	42.229
8×8	2×2	7	6.066	42.118

Table 5.14: Performance of our proposed scheme for the image sequence "Salesman".

maximum block size	minimum block size	Threshold value	Total bit [Kbits]	WPSNR [dB]
8×8	2×2	3	18.851	39.467
8×8	2×2	4	15.960	39.229
8×8	2×2	5	14.098	38.997
8×8	2×2	6	12.755	38.491
8×8	2×2	7	11.936	38.399

Table 5.15: Performance of our proposed scheme for the image sequence "Band".

maximum block size	minimum block size	Threshold value	Total bit [Kbits]	WPSNR [dB]
8×8	2×2	3	14.748	41.008
8×8	2×2	4	11.373	40.963
8×8	2×2	5	8.985	40.717
8×8	2×2	6	7.518	40.323
8×8	2×2	7	6.691	39.660

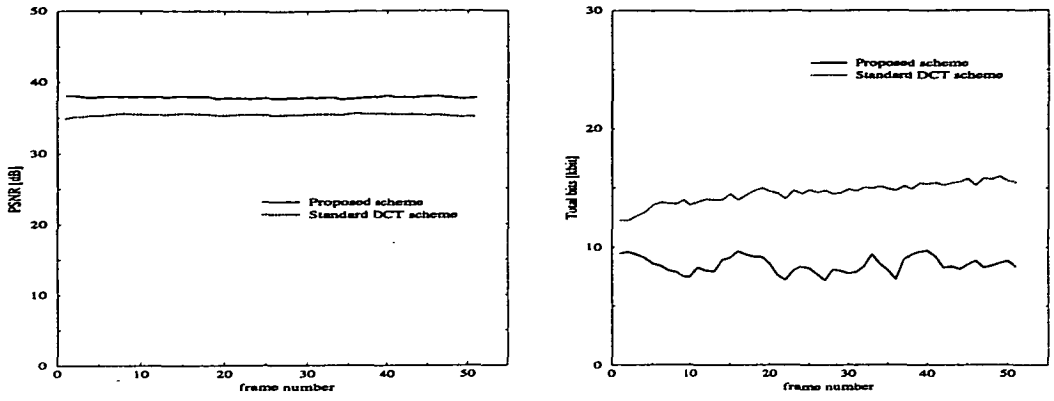
Table 5.16: Performance of our proposed scheme for the image sequence "Skiing".

maximum block size	minimum block size	Threshold value	Total bit [Kbits]	WPSNR [dB]
8×8	2×2	3	27.178	41.238
8×8	2×2	4	23.716	40.614
8×8	2×2	5	21.463	40.111
8×8	2×2	6	19.214	39.590
8×8	2×2	7	17.722	39.117

Table 5.17: Performance of our proposed scheme for the image sequence "Scenic view".

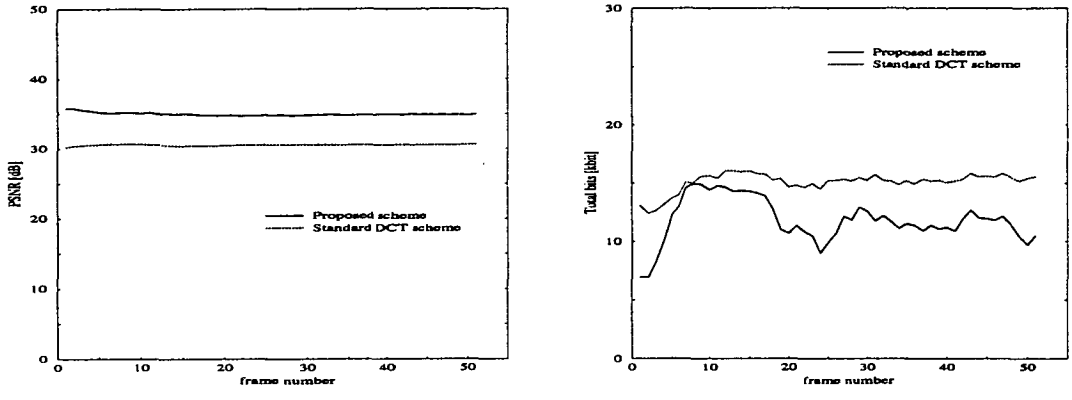
The total number of bits in the columns of Tables i5.11 - 5.17 is the sum of the number of total bits of displacement vectors, the bits of the number of subblocks and the bits of the difference images. The simulation results for all Tables are only carried out for the image sequences of the first and second frames. The simulation results for more frame sequences of the images "Miss America" and "Salesman" are shown in Figure 5.12 (a) and (b), and Figure 5.13 (a) and (b) respectively. The decoding practically is not of interest in this study.

The simulation results of the proposed scheme clearly show that the number of total bits transmitted is less than for the existing schemes for almost the same value of WPSNR. This means that the block matching can be performed in the transform domain without any sacrifice in the quality of the reconstructed image. Moreover, the transform coefficients of the wavelet decomposition can be subjectively thresholded using the HVS before the matching is done resulting in a lesser number of blocks to match and hence a reduction in the number of total bits to be transmitted. The existing scheme when used with the wavelet transform also results in a lesser number of total bits compared to the existing scheme using the DCT. Since the proposed scheme incorporates thresholding, where the threshold value varies, then the number of total bits can also vary. For a threshold value of 3 the increase in performance of the proposed scheme over the existing scheme using the wavelet transform is not very substantial, with decreases in the total number of bits ranging from only 0.4 to 4.7 Kbits for the 5 test images. With respect to the existing scheme using the DCT, however, the decrease in the total number of bits obtained from using the proposed scheme becomes more substantial. These decreases in the number of bits come without



(a) (b)

Figure 5.12: Comparison of the proposed scheme and the standard DCT scheme for image sequences "Miss America", (a) in terms of PSNR and (b) in terms of total bits.



(a) (b)

Figure 5.13: Comparison of the proposed scheme and the standard DCT scheme for image sequences "Salesman", (a) in terms of PSNR and (b) in terms of total bits.

any substantial decrease or change in the WPSNR compared to the existing schemes for the reconstructed images thereby demonstrating the usefulness and superiority of the proposed scheme over the existing schemes. Furthermore, when using thresholds greater than 3 the number of bits is drastically reduced in the proposed scheme with no corresponding decrease in the WPSNR observed, thereby indicating the proposed scheme's insensitivity to the threshold level. The comparative computational cost of the proposed and the standard DCT schemes is 79.147 and 284.238 seconds respectively for the test image sequence "Miss America" on Sun SPARC station IPX operating under conditions of no load. The proposed scheme has a lower computational cost because the block matching is performed in the transform domain, after psychovisual

thresholding. Moreover, the block matching uses a variable block size. This has the advantage that small blocks can be used in areas where the motion is complicated, and thus the boundaries of moving can be effectively matched.

5.5 Summary

The block matching motion compensation scheme is a useful technique for reducing the total number of bits for transmission in video coding. The performance of this scheme depends on the choice of block size and the maximum window size for performing the matching. The simulation results showed that the use of small block sizes resulted in a higher PSNR compared to the PSNR obtained from larger block sizes. Large window sizes also resulted in a higher PSNR, but also resulted in a longer time to find the best match. The block size of 8×8 resulted in the optimum number of total bits for transmission with a maximum window size of ± 8 pixels in both the horizontal and vertical directions. The window size of ± 8 pixels per frame is satisfactory for use in image sequences with frame rates of 25 frames per second allowing low computational times, and giving an adequate PSNR. Due to the variable nature of the images the use of a fixed block size was deemed inappropriate and a variable block size motion compensation algorithm was implemented. Large block sizes are useful for compensating blocks of an image that contain uniform motion or a low level of activity, thereby resulting in a lesser number of total bits to represent the lesser number of displacement vectors. Small block sizes are useful for compensating blocks of an image which have many objects moving at different directions or velocities, thereby resulting in a better quality reconstructed image. The use of a variable block size demonstrated a superior performance compared to the use of a fixed block size in block matching motion compensation. The maximum block size of 8×8 and minimum block size of 2×2 with a window size of ± 8 pixels resulted in the optimum number of total bits for transmission.

In order to reduce further the total of number bits for transmission using block matching motion compensation a combination of motion compensation and transform coding was used. The existing motion compensation and coding schemes employ transform coding using the DCT after motion compensation. Our proposed scheme [5] performs the motion compensation after the transform coding which is implemented using the

wavelet transform. The proposed scheme allows the transform coefficients to be subjectively thresholded before variable block size motion compensation is carried out. Since the wavelet decomposition is a multiresolution technique it is easily adapted for use with the HVS. Therefore, the variable block size motion compensation is carried out psychovisually in the transform domain. The proposed scheme has the advantages that more transform coefficients can be discarded by the subjective thresholding and hence less blocks are matched. The simulation results clearly show that the proposed scheme incorporating the HVS is superior compared to the existing schemes using either the DCT or the wavelet transform. Moreover, in the proposed scheme, the total bit rate can be held constant by varying the quality of the reconstructed image.

Chapter 6

Summary and Future Extensions

6.1 Introduction

This thesis has been concerned with psychovisual image compression techniques using the wavelet transform for image and video coding. Psychovisual image and video coding schemes using subjective thresholding of wavelet transform coefficients have been proposed and implemented. This chapter presents a summary of the final results obtained during the course of this study, as well as outlining a number of possible future extensions.

6.2 An Overview

Applications of image data compression are primarily in the transmission and storage of information. Image compression techniques can be divided into two schemes. These are image compression techniques and video compression techniques. The image compression techniques exploit the redundancies in images a single frame at a time (i.e. intraframe coding), while the video compression techniques exploit the temporal correlations between successive images in the sequence (i.e. interframe coding). If there exists no correlation within the various images making up the sequence, the problem of video compression reduces to one of image compression, with each individual image

within the sequence treated separately. However, if the sequence is highly correlated, the video compression technique holds the promise of significantly greater compression in the data when transmitting an image sequence.

In a number of applications for transmitting video sequences, motion compensation has been successfully applied to video coding. The most common motion compensation technique is the Block Matching Algorithm (BMA). In order to obtain a high rate of data compression, a combination of motion compensated interframe prediction and intraframe DCT encoding has already been implemented and standardized. This technique, however, results in a blocking effect for low bit rates.

The video signals usually contain a considerable amount of statistically and subjectively superfluous information. The video signals can be compressed by using either statistical or psychovisual techniques. A statistical image compression technique exploits statistical redundancies in the information in the image. This technique reduces the amount of data to be transmitted or to be stored in an image with little or information being lost. The aim of psychovisual image compression is to discard the subjective redundancies in an image. This technique relies on discarding information that a human observer is unlikely to perceive and is based on the actual behaviour of the HVS. This technique also offers much greater levels of compression, but it is no longer possible to reconstruct the original image perfectly.

Descriptions of various statistical and psychovisual image coding and compression techniques including a wide range of methods ranging from simple PCM and predictive coding to more complex algorithms involving transform coding, pyramid coding, subband coding and the exploitation of visual psychophysics have been provided. By considering the complexities and compression ratios of these techniques, it was concluded that a combination of statistical transform coding using the wavelet transform and exploitation of visual psychophysics should be used in order to produce an optimal compression algorithm for visual images.

6.3 Summary of Results

The psychovisual image compression of the wavelet transform coefficients in the implemented scheme removes both statistically and psychovisually redundant information from the image during the image coding process by combining the wavelet transform with the characteristics of the HVS. The results obtained using the wavelet transform depend on the choice of wavelet basis. The wavelet basis functions are obtained from a mother wavelet by dilation and translation. These basis function are characterised by a number of parameters, often mutually exclusive. These are spatial compactness, orthonormality, regularity or smoothness and symmetry or anti-symmetry. The wavelet bases can be constructed so as to be either orthonormal or biorthonormal. The family of orthonormal wavelets can be divided into two families, both of them compactly supported, i.e. Daubechies wavelets and coiflets. The Daubechies wavelets are constructed such that there is a unique solution and a maximum number of vanishing moments for the wavelet function only. The coiflets are constructed such that they have a maximum number of vanishing moments for both the scaling and wavelet functions. The construction of biorthonormal wavelets results in symmetrical wavelet bases. The scaling functions of biorthonormal wavelets are the B-splines, and the wavelets are also spline functions. The biorthonormal wavelet bases can also be constructed such that they are close to orthonormal, for example the Laplacian wavelets.

6.3.1 The Image Compression Scheme

The image compression scheme based on the psychovisual thresholding of wavelet transform coefficients using the approach based directly on Mallat's multiresolution wavelet decomposition [18] [69] resulted in a higher WPSNR compared to the statistical compression technique using the wavelet transform for the same bit rates. This result was obtained for both the orthonormal wavelet bases as well as the biorthonormal wavelet bases. The use of the B-spline biorthonormal wavelet with $L = 4$ and $\tilde{L} = 2$ resulted in the highest WPSNR compared to all other wavelet bases considered. Since, in psychovisual image compression, the subjective thresholding and quantization are only applied on the analysis side, only the wavelet bases on the analysis side are taken into account in the HVS. Consequently, when using biorthonormal wavelets where the analysis and synthesis filters are of different length, the less regular wavelet or shorter filter should

correspond to the analysis side since the HVS de-emphasizes the high frequency components. The longer filter should be used on the synthesis side in order to produce a higher quality reconstructed image.

The compression performance of the proposed scheme was evaluated by employing five standard images of size $512 \times 512 \times 8$ bits. Compression ratios varying from $19 : 1$ to $33 : 1$ were obtained for the test images with no observable difference between the original and reconstructed images when viewed from the standard viewing distance. This result was based on a subjective assessment performed by human observers.

The proposed and implemented scheme based on the use of Mallat's algorithm [18], however, resulted in a slightly lesser WPSNR compared to another proposed scheme using the best basis approach with the same constant bit rate. In the best basis approach it is possible to choose the most appropriate basis for representing the given decomposed image using a minimum cost function, while in the conventional approach based on Mallat's algorithm the wavelet bases are fixed. In terms of computational cost, however, the best basis approach is much more expensive. Due to the large computational cost that renders the best basis algorithm impractical for real-time applications, at least with today's computer processor speeds, it was concluded that the proposed subjective wavelet transform coding technique using an approach based on the use of Mallat's algorithm is able to provide an ideal basis for efficient, general purpose video and image compression.

6.3.2 The Video Compression Scheme

The conventional motion compensation scheme using a fixed and variable block matching algorithms was implemented for test image sequences. The simulation results showed that motion compensation using the variable block matching algorithm was superior compared to the fixed block matching algorithm. Consequently the implemented multiresolution motion compensation scheme utilised a variable block matching algorithm. In this scheme the block matching was performed after the subjective thresholding of the wavelet coefficients. This is in contrast to the existing schemes which apply the block matching before the transform coding (either the DCT or wavelet transform). Since the subjective thresholding of the wavelet transform coefficients results in more transform coefficients being discarded, the matching can be performed faster.

The simulation results showed that the proposed scheme is superior in reducing the number of total bits compared to the existing schemes which use either the DCT or the wavelet transform. In the proposed scheme a variation of the given threshold value was employed and due to the nature of the psychovisual-based portion, the total bit rate can be adjusted to produce any desired level of reconstructed image quality.

6.4 Future Extensions

Although the implemented image and video coding algorithms produce high compression ratios and good quality reconstructed images, there is inevitably space for improvements and extensions within this area of research.

In the proposed scheme, all the 8-bit test images have been encoded using only the gray scale or luminance component information in an image. Clearly, for many applications, it would be an advantage to extend the proposed scheme to include colour images. Colour images contain both luminance and chrominance information, and it is well known that the human eye, in general, is much less sensitive to the chrominance component, as opposed to the luminance information [14] [136]. This fact allows noticeably less data to be used to represent the chrominance information in an image [11, 245-299]. Therefore the chrominance information would have to be encoded separately and by exploiting the subjective response of the HVS to stimuli of varying levels of chrominance, an improvement in the compression ratio could be achieved when using our scheme.

There is also the opportunity for improvement in the area of motion compensation in video signals. In particular, motion in images is not limited to translational motion only, but can contain rotation and scaling. In many instances objects can move behind other objects resulting in what is termed occlusion. These effects significantly increase the difficulty in finding the best match in the block matching algorithms. Another consideration in motion compensation that should be explored is the effect of lighting. Variable lighting conditions can change the intensity of the objects in the scene thereby creating problems in accurately matching blocks within an image. Further research is also required to determine suitable parameters for temporal masking, since the subjective perception of temporally changing stimuli is an important consideration. The

frequency sensitivity of the HVS is also greatly reduced over the entire range of the spatial frequencies as the temporal frequency approaches that of motion video signals [11, pp. 245-274]. Therefore by using this type of effect in our scheme the bit rates could be reduced without compromising the perceived quality of the reconstructed image sequence.

6.5 Concluding Remarks

Psychovisual image and video coding algorithms using the wavelet transform in conjunction with the HVS have been developed and implemented. The use of biorthonormal wavelet bases, especially B-Spline wavelet bases, resulted in the highest compression ratios and highest quality reconstructed images. The application of the scheme for image sequences also demonstrated that a greater reduction in the bit rates could be achieved without any degradation in the reconstructed image. The scheme allows the transform coefficients to be subjectively thresholded before variable block size motion compensation is carried out. Since the wavelet decomposition is a multiresolution technique it is easily adapted for use with the HVS. The variable block size motion compensation is carried out psychovisually in the transform domain. The scheme has the advantage that more transform coefficients can be discarded by the subjective thresholding and hence less blocks need to be matched.

Although the scheme resulted in high compression ratios and high quality reconstructed images, and many of the significant considerations relevant to the efficient coding and compression for visual communications have been addressed in this thesis, this research area has not been exhausted. A number of possible extensions and improvements have been suggested in the area of this work. In particular, extensions of this work into colour image compression and video coding would prove useful for many visual communication applications.

Appendix A

Test Images and Sequences

The standard test images used in the image compression study are presented in this appendix. Figures A.1, A.2, A.3, A.4 and A.5 show the five standard test images which have dimensions of size $512 \times 512 \times 8$-bits. They are called "Lenna", "Airplane", "Bird", "Peppers" and "Zelda" respectively.

Figures A.6, A.7 A.8, A.9 and A.10 show the first frame of the test image sequences which have dimensions of size 256×256 with frame rates of 25 frames per second. They are called "Miss America", "Salesman", "Band", "Skiing" and "Scenic view" respectively.

Figure A.1: The standard single frame test image "Lenna".

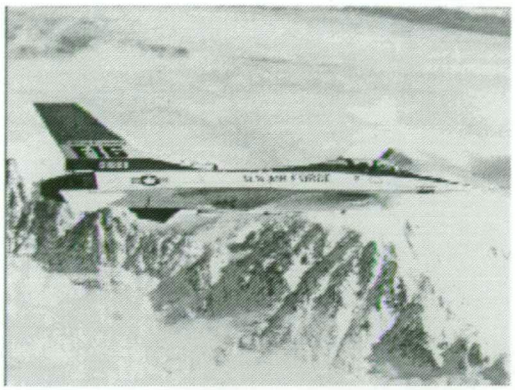


Figure A.2: The standard single frame test image "Airplane"

Figure A.3: The standard single frame test image "Bird".

Figure A.4: The standard single frame test image "Peppers".

Figure A.5: The standard single frame test image "Zelda".

Figure A.6: The first frame of the standard test image sequence "Miss America".

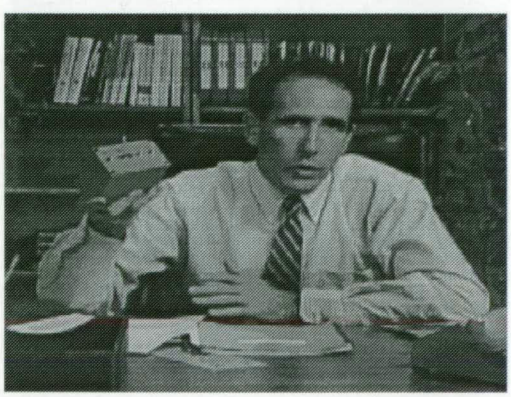


Figure A.7: The first frame of the standard test image sequence "Salesman".

Figure A.8: The first frame of the standard test image sequence "Band".

Figure A.9: The first frame of the standard test image sequence "Skiing".

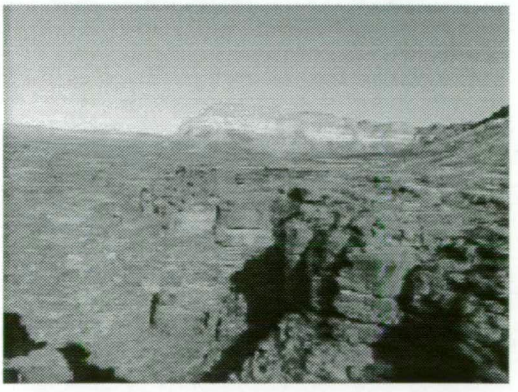


Figure A.10: The first frame of the standard test image sequence "Scenic view".

Appendix B

A Subjective Image Quality Measure

Subjective assessment tests are widely used to evaluate the picture quality of coded images [112]. Since the perceived image quality depends on the subjective nature of a human observer, it is very difficult to determine a distortion criterion appropriate to the human observer. In many circumstances the perceived image quality as observed by a human is strongly dependent on a number of variable factors including the observers background, his/her motivation and the instructions given to the observer [24] [137]. This appendix presents a detailed description of the subjective image quality measure employed in this study.

B.1 Measurement Criterion

The purpose of the subjective measure used in this study is to measure the distortion level resulting from the psychovisual image compression of an image, as adopted in [54]. This measurement is defined as the level that the observer can just distinguish that the original and reconstructed images are different or a degradation in quality is just noticeable in the reconstructed image. Although two observers might both agree that the quality of a reconstructed image using compression technique 1 was worse than

the quality of the reconstructed image using compression technique 2, it is possible that one of the observers might say that the quality of the reconstructed image using compression technique 2 is three times worse and the other observer might say that it is only two times worse. This highlights the difficulties present when using a subjective assessment criterion for image quality measurements.

B.2 Viewing Conditions

Visual distortion in an image will appear different to an observer depending on the particular viewing conditions at the time. In order to improve the objectiveness of the measure, these conditions must be fully defined and maintained throughout the measurement process [54]. Figure B.1 shows the position of the viewing conditions for the subjective measure and Figure B.2 shows the on screen placement of the original and reconstructed images used in this study.

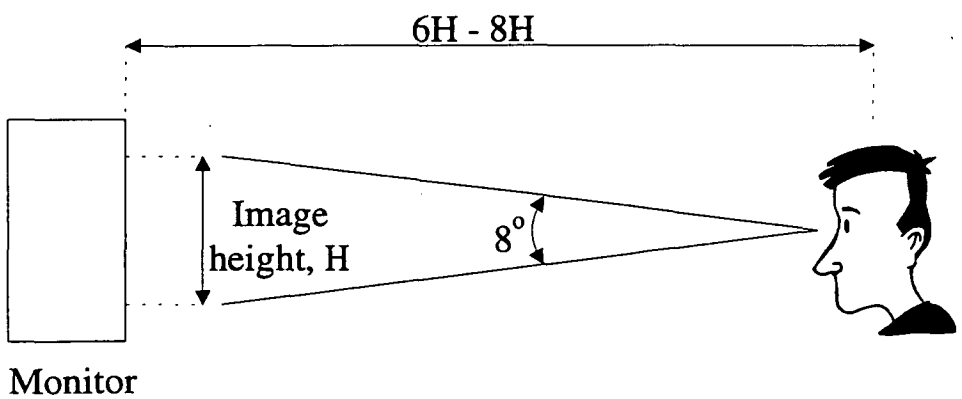


Figure B.1: The viewing conditions.

The images are displayed at broadcast quality on a 20 inch Sun colour monitor whose vertical placement is maintained at eye-level. The distance between the viewer and monitor is given by the international viewing standard [11, pp. 119] [112] of six to eight times the picture height, thereby resulting in a viewing angle of approximately 8 degrees subtended by the viewer as illustrated in Figure B.1. In order to produce the optimum image quality measurement, the brightness and the contrast of the images are maximized. This viewing condition is applied throughout the testing performed in Chapter 4.

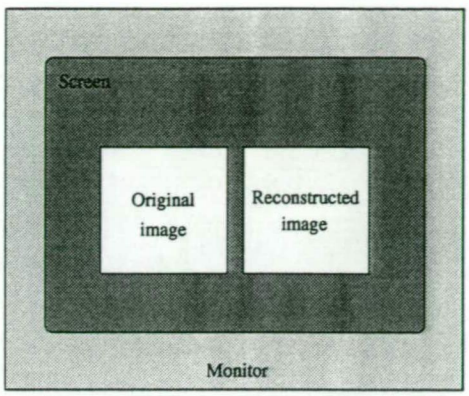


Figure B.2: The placement of the original and reconstructed images.

B.3 The Testing Method

For a particular image compression technique the following process is employed in order to obtain a measurement of the compression level:

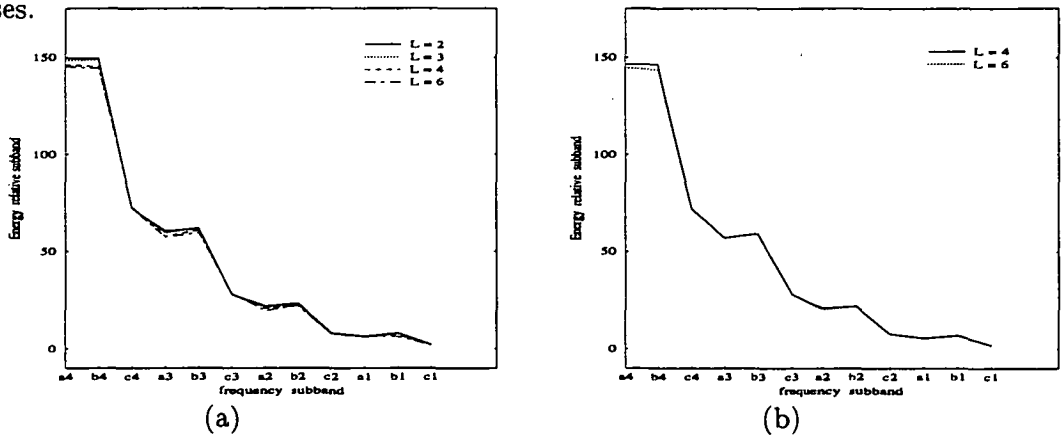
- The original and reconstructed images are displayed on the monitor under the conditions described in Section B.2.
- The viewer then compares both the original and reconstructed images until satisfied that there is no perceivable difference between them.
- This process is repeated using progressively higher compression levels with the same image until the viewer is able to notice some degradation in the reconstructed image.

In order to reduce the subjective nature of the testing method, this process is repeated by using a number of different human viewers. The average distortion compression level is then determined as the level at which 50 % of the viewers begin to see some distortion in the reconstructed image. This method of subjectively measuring image quality was used to determine the optimum parameters for the image compression techniques developed in Chapter 4.

Appendix C

Average Energy Distributions

This appendix contains the average energy distributions of twenty standard test images for various wavelet bases. The frequency subbands shown in the Figures, which are denoted by $a4$, $a3$, $a2$, $a1$, $b4$, $b3$, $b2$, $b1$, $c4$, $c3$, $c2$, and $c1$ on the horizontal axis, correspond to the subbands W_4^{GH}, W_3^{GH}, W_2^{GH}, W_1^{GH}, W_4^{HG}, W_3^{HG}, W_2^{HG}, W_1^{HG}, W_4^{GG}, W_3^{GG}, W_2^{GG}, and W_1^{GG} respectively. Since it is difficult to differentiate between the average energy subbands of each wavelet basis in the Figures, and in order highlight the difference between the average subband energies of the various wavelet bases, this appendix also tabulates the average energy distributions for the various the wavelet bases.



(a) (b)

Figure C.1: The average energy distribution of twenty standard test images using orthonormal wavelet bases. (a) Daubechies wavelets and (b) near linear phase wavelets.

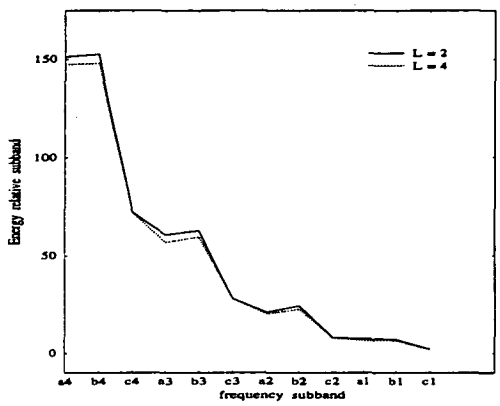


Figure C.2: The average energy distribution of twenty standard test images using orthonormal wavelet bases for coiflet wavelets.

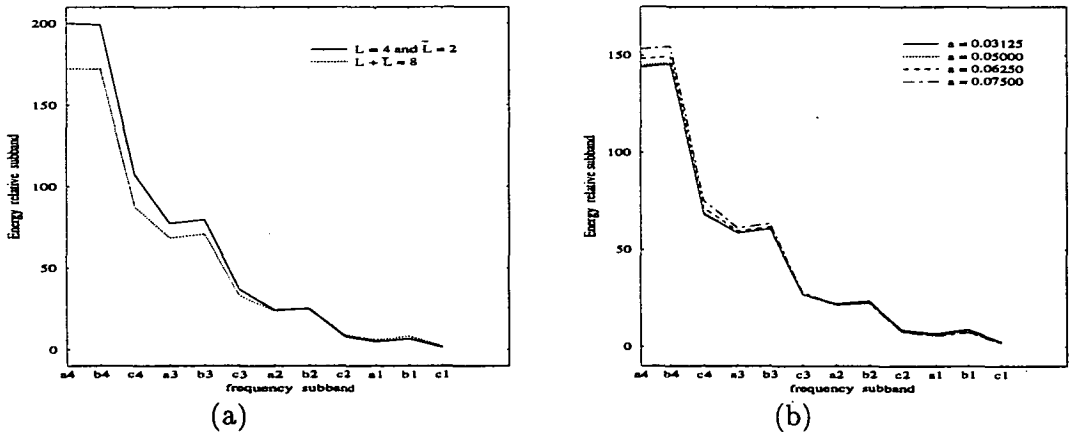


Figure C.3: The average energy distribution of twenty standard test images using biorthonormal wavelet bases. (a) Spline wavelets and (b) Laplacian wavelets.

subband	Average energy $L=2$	Average energy $L=3$	Average energy $L=4$	Average energy $L=6$
V_4^{HH}	1996.274000	1996.924000	1997.940000	1998.653000
W_4^{GH}	149.457200	148.258600	146.188500	145.180700
W_4^{HG}	149.523500	148.732900	145.203600	144.425600
W_4^{GG}	72.151200	72.595950	72.755650	72.304160
W_3^{GH}	60.471650	59.512360	57.750210	57.403990
W_3^{HG}	61.815540	62.257980	61.050500	59.637690
W_3^{GG}	27.965630	27.887570	27.830910	28.154840
W_2^{GH}	21.899480	21.395090	20.743070	19.576200
W_2^{HG}	23.358880	22.533830	22.108020	22.244830
W_2^{GG}	8.263069	7.943480	7.781223	7.556639
W_1^{GH}	6.367443	6.047582	6.299194	6.353568
W_1^{HG}	8.133063	7.692980	7.156462	6.289573
W_1^{GG}	2.203811	2.091296	2.092047	1.974773

Table C.1: The average energy in the subbands for the orthonormal Daubechies wavelet bases.

subband	Average energy $L=4$	Average energy $L=6$
V_4^{HH}	1998.199000	1998.753000
W_4^{GH}	146.537000	144.877100
W_4^{HG}	146.394300	143.569800
W_4^{GG}	72.048440	71.837930
W_3^{GH}	57.057550	57.147660
W_3^{HG}	59.391120	59.411050
W_3^{GG}	28.034960	27.989910
W_2^{GH}	21.071470	20.354930
W_2^{HG}	22.049120	22.023730
W_2^{GG}	7.781597	7.678472
W_1^{GH}	5.674975	5.567358
W_1^{HG}	7.147258	6.843588
W_1^{GG}	2.001996	1.954394

Table C.2: The average energy in the subbands for the orthonormal near linear phase wavelet bases.

	Average energy	Average energy
subband	$L = 2$	$L = 4$
V_4^{HH}	1995.641000	1998.019000
W_4^{GH}	151.354800	147.395800
W_4^{HG}	152.640500	148.017100
W_4^{GG}	72.595250	72.201900
W_3^{GH}	60.727200	56.987050
W_3^{HG}	62.870630	59.633680
W_3^{GG}	28.053130	28.014620
W_2^{GH}	21.142970	20.262460
W_2^{HG}	24.241650	22.626790
W_2^{GG}	8.114229	7.738266
W_1^{GH}	7.535518	6.625604
W_1^{HG}	7.160048	6.416955
W_1^{GG}	2.205952	2.009508

Table C.3: The average energy in the subbands for the orthonormal Coiflet wavelet bases.

	Average energy	Average energy
subband	$\tilde{L} = 2\ L = 4$	$\tilde{L} + L = 8$
V_4^{HH}	2031.511000	2007.132000
W_4^{GH}	200.016600	172.237600
W_4^{HG}	199.158500	171.948700
W_4^{GG}	107.268800	87.482900
W_3^{GH}	77.397650	68.487480
W_3^{HG}	79.588660	70.791680
W_3^{GG}	37.214500	33.371750
W_2^{GH}	24.284830	23.654850
W_2^{HG}	24.902350	25.307330
W_2^{GG}	8.304849	9.180062
W_1^{GH}	5.051787	6.184390
W_1^{HG}	6.603382	8.117710
W_1^{GG}	1.424274	2.215493

Table C.4: The average energy in the subbands for the biorthonormal Spline wavelet bases.

	Average energy	Average energy	Average energy	Average energy
subband	$a = 0.03125$	$a = 0.05000$	$a = 0.06250$	$a = 0.07500$
V_4^{HH}	1987.116000	1992.641000	1996.828000	2001.496000
W_4^{GH}	144.062500	145.060800	148.338100	153.471300
W_4^{HG}	145.496800	146.283800	149.400300	154.368400
W_4^{GG}	68.245480	69.160760	71.640780	75.378960
W_3^{GH}	58.837730	58.776230	59.740650	61.393820
W_3^{HG}	61.018370	61.018470	61.989740	63.622170
W_3^{GG}	27.027940	26.786200	27.255050	28.117330
W_2^{GH}	22.223820	21.601000	21.501600	21.598060
W_2^{HG}	23.583450	22.757890	22.523820	22.485700
W_2^{GG}	8.701665	8.045763	7.811048	7.688025
W_1^{GH}	6.815205	6.174414	5.858797	5.609244
W_1^{HG}	8.996346	8.178728	7.760748	7.419806
W_1^{GG}	2.587443	2.187527	1.992147	1.834922

Table C.5: The average energy in the subbands for the biorthonormal close to orthonormal wavelet bases (Laplacian wavelet bases).

Appendix D

Tables of Huffman Code Words

This appendix describes the distribution of transform coefficients for the standard test images as well as presenting the table of Huffman code words used for coding the bit streams. The Huffman code words are generated from the average distribution of quantized transform coefficients by using the method of minimum redundancy code construction [11, pp. 148] [32]. The Huffman code words consist of the run length, the combination of coefficients amplitudes, Beginning Of Subblock (BOS) markers and Run-Length Prefix (RLP) markers. Code words for coefficients of magnitude between -15 and $+15$, and run length from 1 to 64 have been determined. Any coefficient magnitudes or run lengths which exceed these values are simply encoded by the Huffman code for this maximum value followed by an 8-bit binary code word which represents the actual value. The activity function, A_F, is also coded using an additional 8-bit binary code word. This does not unduly affect the average bit rate since the activity function is coded only once every subband decomposition. Figures D.1 and D.2 show the distributions of transform coefficients for the standard test images. The Huffman code words for the transform coefficients and run lengths are tabulated in Tables D.1 and D.2.

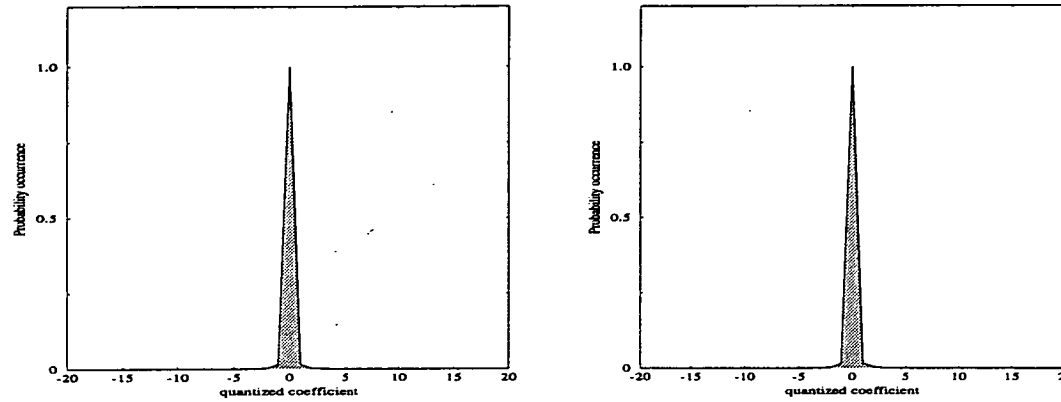


Figure D.1: The quantized coefficient distributions for the standard test images. (a) "Lenna" image and (b) "Airplane" image.

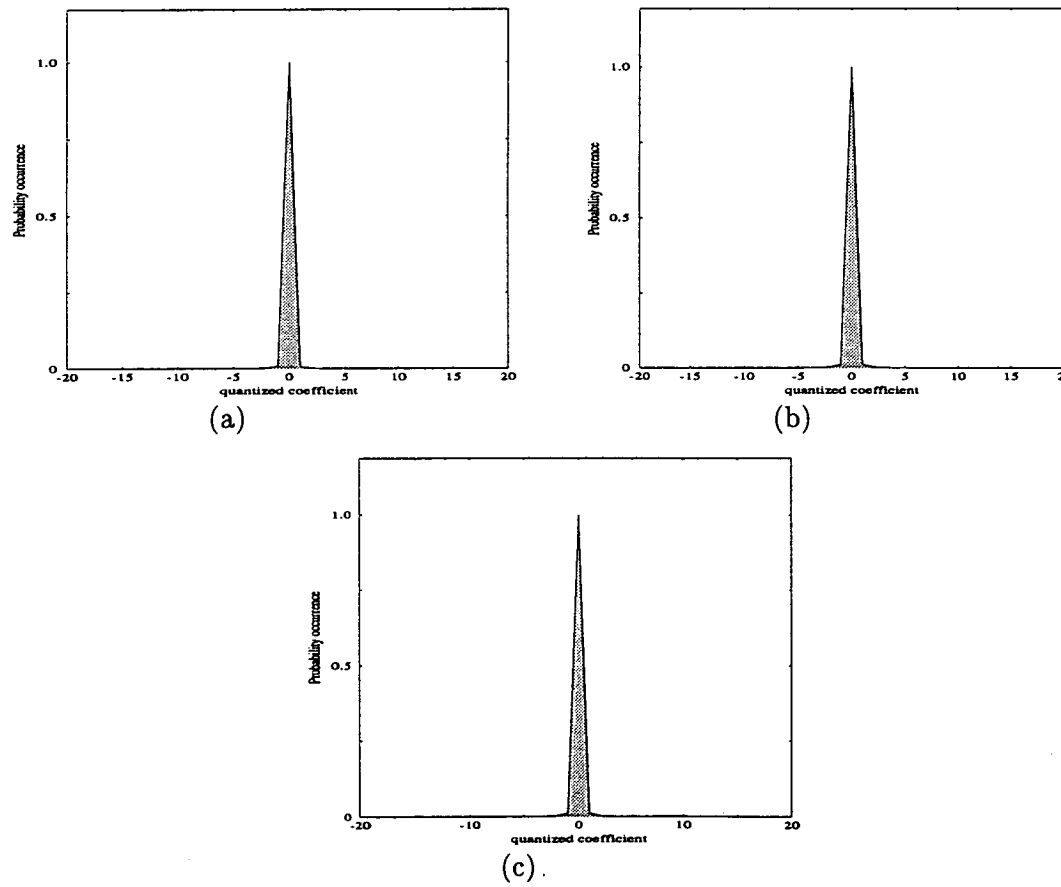


Figure D.2: The quantized coefficient distributions for the standard test images. (a) "Bird" image, (b) "Peppers" image and (c) "Zelda" image.

Amplitude value	Probability	Huffman Code Word			
$\leq$ -15	0.0001	1110	0011	0000	1 + 8
-14	0.0001	1110	0011	0000	00
-13	0.0003	1110	1011	100	
-12	0.0005	1110	1011	01	
-11	0.0002	1110	0011	0001	
-10	0.0004	1110	0011	001	
-9	0.0004	1110	1011	11	
-8	0.0012	1110	0011	1	
-7	0.0025	1110	1010		
-6	0.0046	1110	0000		
-5	0.0096	1110	01		
-4	0.0147	1001	10		
-3	0.0299	1001	0		
-2	0.0711	1000			
-1	0.3145	01			
RLP	0.0895	101			
BOS	0.0027	1110	0010		
1	0.3218	00			
2	0.0730	110			
3	0.0306	1111			
4	0.0140	1001	11		
5	0.0082	1110	11		
6	0.0048	1110	100		
7	0.0018	1110	0001	0	
8	0.0016	1110	0001	1	
9	0.0007	1110	0011	01	
10	0.0002	1110	1011	0010	
11	0.0004	1110	1011	000	
12	0.0001	1000	1100	0001	
13	0.0001	1110	1011	0011	0
14	0.0003	1110	1011	101	
$\geq$ 15	0.0001	1101	0110	0111	+ 8
Average Bits per Amplitude Code: 2.79					

Table D.1: The Huffman code words for the transform coefficients.

Amp. value	Prob.	Huffman Code Word	Amp. value	Prob.	Huffman Code Word
1	0.3753	00	33	0.0011	1111 1001 1
2	0.1937	10	34	0.0008	0111 1100 01
3	0.1032	110	35	0.0008	1111 0100 10
4	0.0666	0100	36	0.0008	1111 0100 11
5	0.0447	1110	37	0.0007	1111 0101 01
6	0.0316	0101 1	38	0.0004	0111 1100 110
7	0.0265	0111 0	39	0.0006	1111 1001 00
8	0.0176	0101 00	40	0.0004	0111 1100 111
9	0.0172	0101 01	41	0.0005	0110 0010 101
10	0.0128	0111 10	42	0.0009	0110 1000 11
11	0.0138	0110 01	43	0.0007	1111 1000 00
12	0.0131	0110 11	44	0.0003	0110 0010 1001
13	0.0059	1111 000	45	0.0005	0110 1000 100
14	0.0059	1111 001	46	0.0006	1111 1001 01
15	0.0065	0110 101	47	0.0008	1111 0100 00
16	0.0056	1111 011	48	0.0009	0111 1100 00
17	0.0062	0111 111	49	0.0002	1111 0100 0110
18	0.0084	0110 000	50	0.0002	1111 0100 0111
19	0.0046	1111 101	51	0.0002	1111 0100 0100
20	0.0020	0110 0011 0	52	0.0003	1111 1000 110
21	0.0042	1111 111	53	0.0004	0111 1100 100
22	0.0045	1111 110	54	0.0001	0111 1100 1011 0
23	0.0018	0110 1001 0	55	0.0001	0111 1100 1011 1
24	0.0019	0110 1000 0	56	0.0005	0110 1000 101
25	0.0031	0111 1101	57	0.0005	0110 0011 110
26	0.0010	0110 0010 11	58	0.0002	1111 0100 0101
27	0.0018	0110 1001 1	59	0.0010	0110 0011 10
28	0.0013	1111 0101 1	60	0.0003	1111 1000 111
29	0.0021	0110 0010 0	61	0.0002	0111 1100 1010
30	0.0006	1111 1000 01	62	0.0003	1111 1000 100
31	0.0003	0110 0010 1000	63	0.0003	1111 1000 101
32	0.0007	1111 0101 00	≥ 64	0.0005	0110 0011 111 + 8
Average Bits per Run-length Code: 3.31					

Table D.2: The Huffman code words for the run length.

Appendix E

A Weighted PSNR Measure

Signal to noise ratio (SNR) is widely used as a measure of the picture quality in image coding schemes. SNR is a simple and tractable measure and can be defined as

$$\text{SNR} = 10\log_{10}\frac{255^2}{E[e^2(x,y)]} \tag{E.1}$$

where $e(x,y)$ is the image coding error, $E[.]$ means the expectation and 255 is the peak value for the 8-bit representation of grey level images [11, pp. 181]. This SNR is more widely known as the peak SNR (PSNR).

The PSNR alone, however, is not an adequate reflection of the quality in terms of subjective assessment. For example, the error sequence $\{10, 10, 10, 10, 10\}$ has the same SNR as $\{0, 0, 10\sqrt{5}, 0, 0\}$, although the spatial distribution of the errors is very different. Since the perceived quality of an image is dependent on the human observer, it is important to take this into account when assessing the quality of the reconstructed image using the PSNR. Much research has been devoted to this aspect of image quality assessment and detailed discussions can be found in References [123] [137] [138] [139] [140].

The subjective assessment method which is used in this study to evaluate the picture quality of reconstructed images is a weighted PSNR (WPSNR). The block diagram for the computation of the WPSNR is illustrated in Figure E.1. The WPSNR is computed as follows : Firstly, the error image is calculated by subtracting the original from the

reconstructed image. Secondly, the error image is Fourier transformed and multiplied by the frequency response of the HVS so that only perceived errors are retained. This is then inverse Fourier transformed to form the weighted error image, $\tilde{e}(x,y)$. The WPSNR is computed from the peak power in the original image and the weighted error image as follows,

$$\text{WPSNR} = 10\log_{10}\left(\frac{255^2}{\text{Power } \tilde{e}(x,y)}\right) \text{dB} \tag{E.2}$$

Figure E.1: The block diagram for computing the WPSNR.

Appendix F

Tables of Displacement Vector Code Words

This appendix contains the one-dimensional code words which are commonly used for coding displacement vectors [8, pp. 133] [132]. These one-dimensional code words are tabulated in Table F.1 on the following page. These code words are used for coding the displacement vectors in Chapter 5.

amplitude displacement vector	code word		
-16 & 16	0000	0011	001
-15 & 17	0000	0011	011
-14 & 18	0000	0011	101
-13 & 19	0000	0011	111
-12 & 20	0000	0100	001
-11 & 21	0000	0100	011
-10 & 22	0000	0100	11
-9 & 23	0000	0101	01
-8 & 24	0000	0101	11
-7 & 25	0000	0111	
-6 & 26	0000	1001	
-5 & 27	0000	1011	
-4 & 28	0000	111	
-3 & 29	0000	1	
-2 & 30	0011		
-1 & 31	011		
0	1		
1 & -31	010		
2 & -30	0010		
3 & -29	0001	0	
4 & -28	0000	110	
5 & -27	0000	1010	
6 & -26	0000	1000	
7 & -25	0000	0110	
8 & -24	0000	0101	10
9 & -23	0000	0101	00
10 & -22	0000	0100	10
11 & -21	0000	0100	010
12 & -20	0000	0100	000
13 & -19	0000	0011	110
14 & -18	0000	0011	100
15 & -17	0000	0011	010

Table F.1: Codes for displacement vectors.

Appendix G

Graphed and Tabulated MAD Values

This appendix contains the plots of the MAD values for each of the test image sequences. The computation of the MAD value is described in Section 5.2. In order to differentiate between the MAD value for each of the test image sequences, this appendix also tabulates the MAD values for each test image sequence and for each subblock.

block size	minimum MAD value	maximum MAD value	average value	standard deviation
2×2	0.0000	20.7500	2.5744	1.1783
4×4	0.4375	17.7500	3.2444	1.4140
8×8	1.3125	13.2500	3.9650	1.4530
16×16	2.3806	10.2700	4.5143	1.2887
32×32	2.9658	8.45210	4.8533	1.0067

Table G.1: The MAD values for the test image sequence "Miss America".

block size	minimum MAD value	maximum MAD value	average value	standard deviation
2×2	0.0000	14.5000	2.9133	1.2140
4×4	0.4375	14.8750	3.9698	1.3508
8×8	1.4688	11.6250	4.7073	1.1403
16×16	2.7148	9.58200	5.0701	0.8777
32×32	4.1279	5.28550	5.2855	0.7421

Table G.2: The MAD values for the test image sequence "Salesman".

block size	minimum MAD value	maximum MAD value	average value	standard deviation
2×2	0.0000	31.5000	3.0317	2.0317
4×4	0.1250	26.6250	4.2744	2.8124
8×8	0.7812	28.8120	5.2258	2.9497
16×16	1.2656	24.5310	5.8781	2.8998
32×32	2.3320	13.5430	6.2423	2.3729

Table G.3: The MAD values for the test image sequence "Band".

block size	minimum MAD value	maximum MAD value	average value	standard deviation
2×2	0.0000	26.5000	2.8842	1.7389
4×4	0.3750	21.7500	3.9926	2.0595
8×8	0.9687	20.5250	4.8282	1.7418
16×16	1.7656	12.0940	5.2933	1.3207
32×32	3.8438	11.7990	5.5863	1.0723

Table G.4: The MAD values for the test image sequence "Skiing".

block size	minimum MAD value	maximum MAD value	average value	standard deviation
2×2	0.0000	26.0000	3.2395	1.7467
4×4	0.5625	21.0620	5.1409	2.5993
8×8	1.8281	18.2660	5.9299	2.5183
16×16	2.8906	15.0700	6.2594	2.2289
32×32	3.3760	15.6340	6.4803	2.2530

Table G.5: The MAD values for the test image sequence "Scenic view".

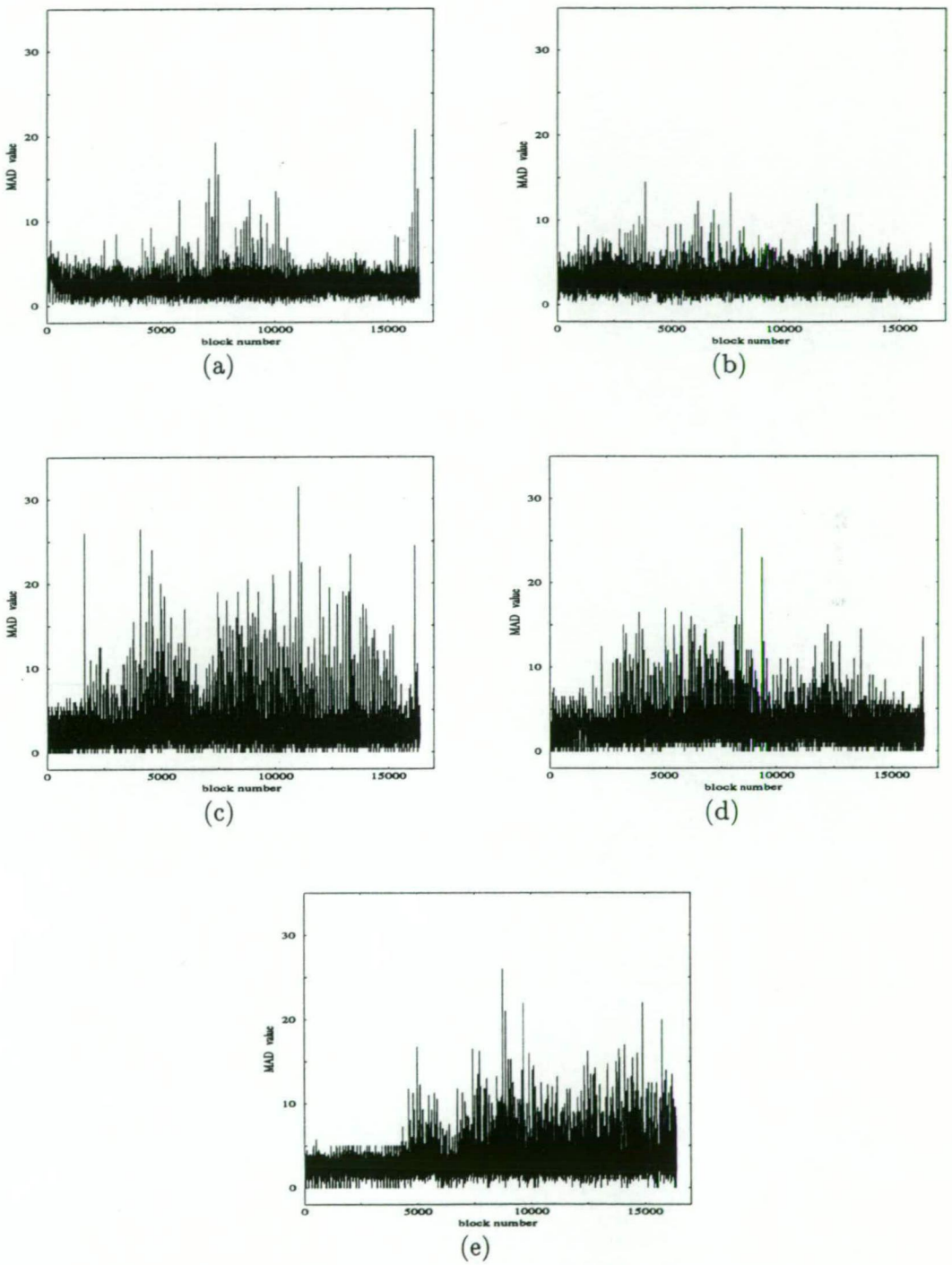


Figure G.1: The MAD value versus block number for the image sequence divided into subblocks of 2×2. *(a) Test image "Miss America", (b) test image "Salesman", (c) test image "Band", (d) test image "Skiing" and (e) test image "Scenic view".*

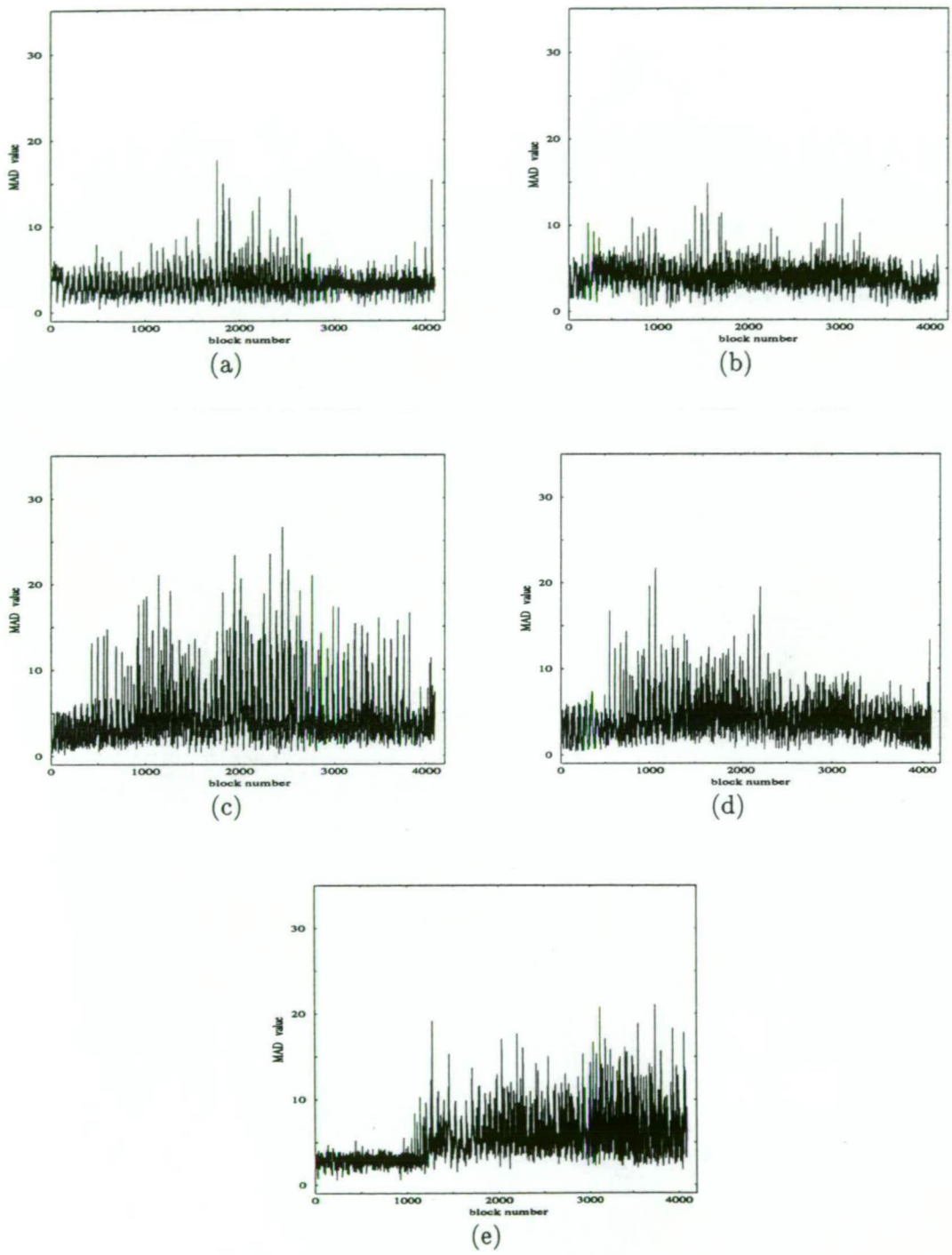


Figure G.2: The MAD value versus block number for the image sequence divided into subblocks of 4×4. *(a) Test image "Miss America", (b) test image "Salesman", (c) test image "Band", (d) test image "Skiing" and (e) test image "Scenic view".*

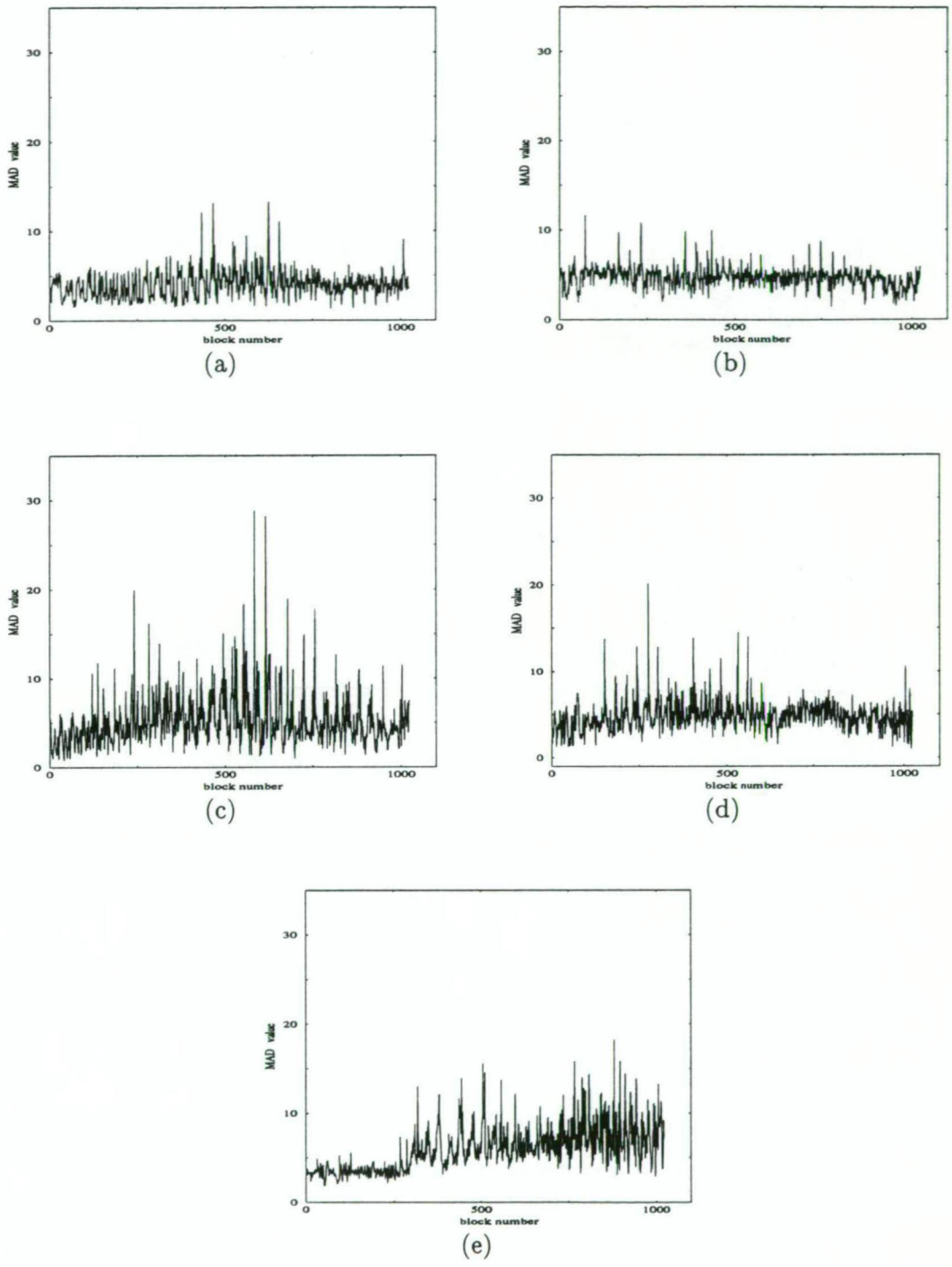


Figure G.3: The MAD value versus block number for the image sequence divided into subblocks of 8×8. *(a) Test image "Miss America", (b) test image "Salesman", (c) test image "Band", (d) test image "Skiing" and (e) test image "Scenic view".*

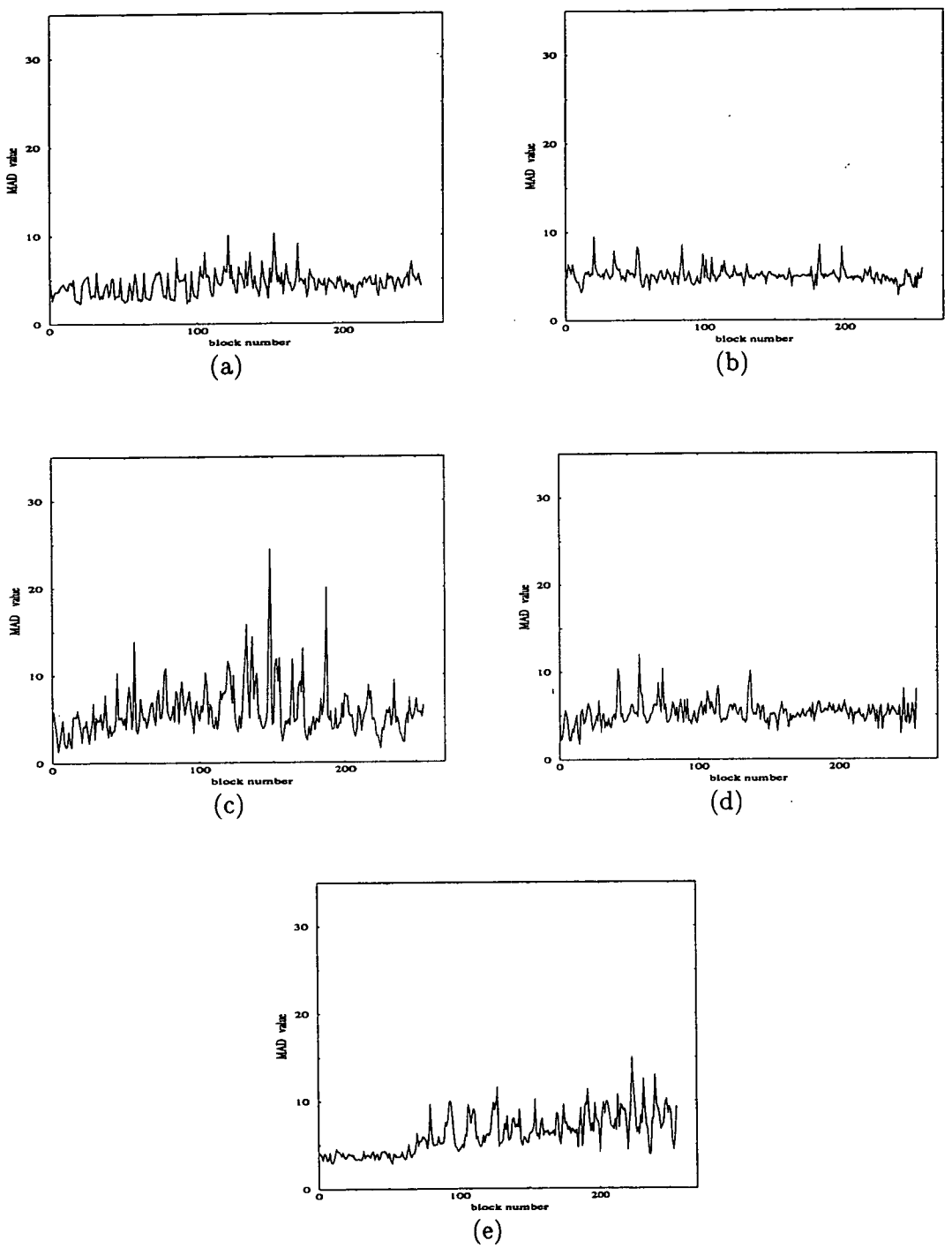


Figure G.4: The MAD value versus block number for the image sequence divided into subblocks of 16 × 16. *(a) Test image "Miss America", (b) test image "Salesman", (c) test image "Band", (d) test image "Skiing" and (e) test image "Scenic view".*

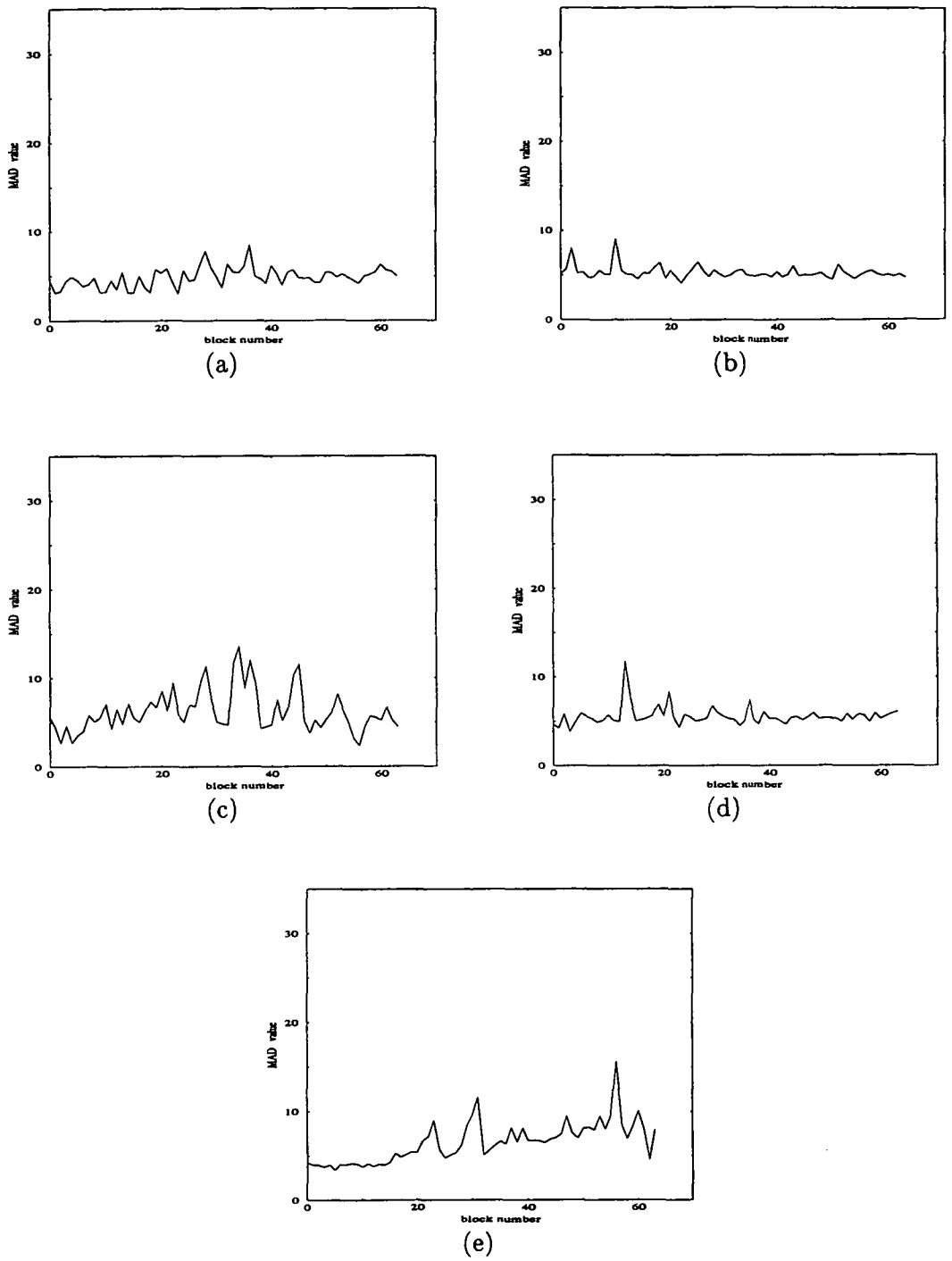


Figure G.5: The MAD value versus block number for the image sequence divided into subblocks of 32×32. *(a) Test image "Miss America", (b) test image "Salesman", (c) test image "Band", (d) test image "Skiing" and (e) test image "Scenic view".*

References

[1] D. T. Nguyen and D. Gunawan, "Wavelets and Wavelets - Design Issues," *1994 IEEE Singapore International Conference on Communication Systems, ICCS'94*, pp. 188–194, Nov. 1994.

[2] D. Gunawan and D.T. Nguyen, "Psychovisual Image Coding Using Wavelet Transform," *Australian Journal of Intelligent Information Processing Systems*, vol. 2, pp. 45–52, Mar. 1995.

[3] D. Gunawan, D.T. Nguyen, and R. G. Lane, "Subjective Coding of Best Basis Wavelet Coefficients for Multiresolution Analysis," *Submitted to Signal Processing: Image Communication*, Dec. 1994.

[4] D. L. McLaren, D. Gunawan and R. G. Lane, "Motion Detection and Adaptive Compensation for Efficient Interframe Video Coding," *The Third International Symposium on Signal Processing and Its Applications (ISSPA'92) Proceedings, Gold Coast, Australia*, pp. 634–637, Aug. 1992.

[5] D. Gunawan, R.G. Lane and D.T. Nguyen, "Adaptive Motion-Compensated Interframe Prediction Prediction Coding Using Subjective Thresholding of Wavelet Transform Coefficients," *1994 IEEE Singapore International Conference on Communication Systems, ICCS'94, Singapore*, pp. 1135–1137, Nov. 1994.

[6] G. Wallace, "The JPEG still picture compression standard," *Communication ACM*, vol. 34, pp. 30–44, Apr. 1991.

[7] Ronald Plompen, *Motion Video Coding for Visual Telephony.* PTT Research Neher Laboratories, Rotterdam, 1989.

[8] Jan Gecsei, *The Architecture of Videotex Systems, pp. 3-13.* Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1983.

[9] A. F. Alber, *Videotex/Teletext, pp. 1-39.* McGraw Hill, New York, 1985.

[10] H. Armbuster, "Applications of Future Broad-Band Services in the Office and Home," *IEEE Communication Magazine*, vol. 4, pp. 429–437, July 1986.

[11] Arun N. Netravali and Barry G. Haskell, *Digital Pictures : Representation and Compression.* Plenum Press, New York, 1988.

[12] K. Habara, "ISDN : A look at the future through the past," *IEEE Communication Magazine*, vol. 26, pp. 25–32, Nov. 1988.

[13] M. Liou, "Visual Telephony as an ISDN Application," *IEEE Communication Magazine*, vol. 28, pp. 30–38, Feb. 1990.

[14] A. Wong, C. T. Chen, D. LeGall, F. C. Jeng and K. M. Uz, "MCPIC : A Video Coding Algorithm for Transmission and Storage Applications," *IEEE Communication Magazine*, vol. 28, pp. 24–32, Nov. 1990.

[15] H. T. Mouftah, "Multimedia Communications : An Overview," *IEEE Communication Magazine*, vol. 30, pp. 18–19, May 1992.

[16] F. Kretz and F. Colaitis, "Standardizing Hypermedia Information Objects," *IEEE Communication Magazine*, vol. 30, pp. 60–70, May 1992.

[17] CCIR Rec. 601., "Digital methods of transmitting television in formation. Recomendation 601, Encoding parameters of digital television for studio," 1982.

[18] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition : The Wavelet Representation," *IEEE Transaction on Patt. Anal. and Mach. Intelligence*, vol. 11, pp. 674–693, July 1989.

[19] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithm for best basis selection," *IEEE Transaction on Information Theory*, vol. 38, pp. 713–718, Mar. 1992.

[20] A. K. Jain, "Image Data Compression," *Proceedings of the IEEE*, vol. 69, pp. 349–389, Mar. 1981.

[21] M. H. Chan, "Image Coding Algorithms for Video-Conferencing Applications," *Ph.D thesis, Imperial College, University of London, London*, 1989.

[22] W. K. Pratt, *Digital Image Processing : Chapter 10 - 11.* John Wiley and Sons Inc., Canada, 1978.

[23] N. S. Jayant and P. Noll, *Digital Coding of Waveforms Principles and Applications to Speech and Video, pp. 510-590.* Prentice-Hall Inc., New Jersey, 1984.

[24] J. O. Limb, "Distortion Criteria of the Human Viewer," *IEEE Transactions on Systems, Man. and Cybernetics*, vol. 9, pp. 778–793, Dec. 1979.

[25] N. Jayant, J. Johnston, and R. Safranek, "Signal Compression Based on Models of Human Perception," *Proceedings of the IEEE*, vol. 81, pp. 1385–1422, Oct. 1993.

[26] C. C. Cutler, "Differential Quantization of Communication Signals," *U.S. Patent 2 605 361*, July 1952.

[27] F. deJager, "Delta Modulation, A Method of PCM Transmission using a 7-unit Code," *Philips Research Reports*, pp. 442–466, Dec. 1952.

[28] C. W. Horrison, "Experiments with linear Prediction in Television," *The Bell-System Technical Journal*, vol. 31, pp. 764–863, July 1952.

[29] H. G. Musmann, "Predictive Image Coding," *in Image Transmission Techniques, W. K. Pratt Edit, Academic Press, New York.*, 1979.

[30] W. H. Chen and W. K. Pratt, "Scene Adaptive Coder," *IEEE Transaction on Communication*, vol. 32, pp. 280–287, Mar. 1984.

[31] R. G. Gallager, "Information Theory and Reliable Communication," *John Willey and Sons, Inc., New York*, 1968.

[32] D. A. Huffman, "A Method for the Construction Minimum Redundancy Codes," *Proceeding IRE*, vol. 40, pp. 1098–1101, Sept. 1952.

[33] J. E. Lim, *Two-dimensional Signal and Image Processing, pp. 589-670.* Prentice Hall, Englewood Cliffs, New Jersey, 1990.

[34] P. A. Wintz, "Transform Picture Coding," *Proceedings of the IEEE*, vol. 60, pp. 809–820, July 1972.

[35] CCITT SG XV WP/1/Q4, "Specialists Group on Coding for Visual Telephony. Description of reference model 8 (rm8)," *Document 525*, June 1989.

[36] CCITT H.261, "Video codec for audiovisual services at p×64kbps," 1989.

[37] M. Ohta and S. Nogaki, "Hybrid Picture Coding with Wavelet Transform and Overlapped Motion-Compensated Interframe Prediction Coding," *IEEE Transaction on ASSP*, vol. 40, pp. 3416–3424, Dec. 1993.

[38] W. K. Pratt, J. Kane and H. C. Andrews, "Hadamard Transform Coding of Images," *Proceedings of the IEEE*, vol. 57, pp. 58–68, Jan. 1969.

[39] N. Ahmad, T. Natarajan and K.R. Rao, "Discrete Cosine Transform Coding of Images," *IEEE Transaction on Computer*, vol. 23, pp. 90–93, June 1974.

[40] W. H. Chen, H. Smith and S. C. Fralick, "A fast computational algorithm for the DCT," *IEEE Transaction on Communication*, vol. 25, pp. 1004–1009, Sept. 1977.

[41] B. G. Lee, "A New Algorithm to compute the DCT," *IEEE Transaction on ASSP*, vol. 32, pp. 1243–1245, June 1984.

[42] N. Ta, Y. Attikiozel and G. Crabbin, "Fast Computation of Two Dimensional DCT using Discrete Radon Transform," *Electronics Letters*, vol. 27, pp. 82–84, Jan. 1991.

[43] P. Goupillaud, A. Grossman and J. Morlet, "Cycle Octave and related transforms in seismic signal analysis," *Geoexploaration*, vol. 23, pp. 85–102, 1984-85.

[44] P. Flandrin, "Wavelet Analysis and Synthesis of Fractal Brownian Motion," *IEEE Transaction on Information Theory*, vol. 38, pp. 910–917, Mar. 1992.

[45] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transaction on Communication*, vol. 31, pp. 532–540, Apr. 1983.

[46] J. W. Wood and S. D. O'Neil, "Subband Coding of Images," *IEEE Transaction on ASSP*, vol. 34, pp. 1278–1288, Oct. 1986.

[47] M. Vetterli, "Multi-dimensional Subband Coding," *Signal Processing*, vol. 6, pp. 97–112, Apr. 1984.

[48] T. N. Cornsweet, *Visual Perception, pp. 276-335.* Academic Press, New York, 1970.

[49] C. F. Hall and E. L. Hall, "A Nonlinear Model for the Spatial Characteristics of the Human Visual System," *IEEE Transactions on Systems, Man. and Cybernetics*, vol. 7, pp. 161–170, Mar. 1977.

[50] A. P. Ginsburg, M. W. Cannon and M. A. Nelson, "Suprathreshold Processing of Complex Visual Stimuli : Evidence for Linearity in Contrast Perception," *Science*, vol. 280, pp. 619–621, May 1980.

[51] F. W. Campbell and D. Green, "Optical and Retina Factors Affecting Visual Resolution," *Journal of Physiology*, vol. 181, pp. 576–593, 1965.

[52] M. D. Levine, *Vision in Man and Machine, pp. 100-150.* McGraw-Hill Book Company, New York, 1985.

[53] R. P. Dooley, "Predicting Brightness Appearance at Edges Using Linear and Non-Linear Visual Describing Functions," *SPIE Annual Meeting, Denver*, May 1975.

[54] David L. McLaren, "Video and Image Coding for Broadband-ISDN," *Ph.D Thesis, University of Tasmania, Tasmania*, 1992.

[55] D. H. Kelly, "Spatial Frequency Selectivity in the Retina," *Vision Research*, vol. 15, pp. 665–672, June 1975.

[56] H. R. Wilson and C. Giese, "Threshold Visibility of Frequency Gradient Patterns," *Vision Research*, vol. 17, pp. 1177–1190, Oct. 1977.

[57] H. R. Wilson, "Quantitative Prediction of Line Spread Function Measurements: Implications for Channel Bandwidths," *Vision Research*, vol. 18, pp. 493–496, Apr. 1978.

[58] K. N. Ngan and K. S. Leong and H. Singh, "Adaptive Cosine Transform Coding of Images in Perceptual Domain," *IEEE Transaction on ASSP*, vol. 37, pp. 1743–1750, Nov. 1989.

[59] B. Chitprasert and K. R. Rao, "Human Visual Weighted Progressive Image Transmission," *IEEE Transaction on Communication*, vol. 38, pp. 1040–1044, July 1990.

[60] A. N. Netravali and P. Birendra, "Adaptive Quantization of Picture Signals using Spatial Masking," *Proceedings of the IEEE*, vol. 65, pp: 536–548, Apr. 1977.

[61] G. E. Legge and J.M. Foley, "Contrast masking in human vision," *Journal of the Optical Society of America*, vol. 70, pp. 1458–1471, Dec. 1980.

[62] A. Fiorentini and M. T. Zoli, "Detection of a Target Superimposed to a Step Pattern of Illumination II. Effects of Just-Perceptible Illumination Step," *Att. Fond. G. Ronchi*, vol. 22, pp. 207–217, 1967.

[63] G. Sperling, "Temporal and Spatial Visual Masking I : Masking by impulse flashes," *Journal of the Optical Society of America*, vol. 55, pp. 541–599, 1964.

[64] C. F. Stromeyer and B. Julesz, "Spatial frequency masking in vision : Critical band and spread of masking," *Journal of the Optical Society of America*, vol. 62, pp. 1221–1232, Oct. 1972.

[65] D. L. McLaren and D. T. Nguyen, "Activity Function for DCT Coded Images," *Electronics Letters*, vol. 25, pp. 1704–1705, Dec. 1989.

[66] D. J. Sakrison and J. L. Mannos, "The Effects of a Visual Fidelity Criterion on the Encoding of Images," *IEEE Transaction on Information Theory*, vol. 20, pp. 525–536, July 1974.

[67] D. L. McLaren and D. T. Nguyen, "The Removal of Subjective Redundancy from DCT Coded Images," *IEE Proceedings - Part I*, vol. 138, pp. 345–350, Oct. 1991.

[68] J. B. Nill, "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Transaction on Communication*, vol. 33, pp. 551–557, June 1985.

[69] S. G. Mallat, "Multifrequency channel decompositions of image and wavelet models," *IEEE Transaction on ASSP*, vol. 37, pp. 2091–2110, Dec. 1989.

[70] O. Riol and M. Vetterli, "Wavelets and Signal Processing," *IEEE Signal Processing Magazine*, vol. 8, pp. 14–38, Oct. 1991.

[71] M. Vetterli and D. Le Gall, "Filter banks allowing perfect reconstruction," *Signal Processing*, vol. 10, pp. 219–244, July 1986.

[72] M. Vetterli and C. Herley, "Wavelets and Filter Banks : Theory and Design," *IEEE Transaction on Signal Processing*, vol. 40, pp. 2207–2232, Sept. 1992.

[73] H. R. Wilson and J.R. Bergen, "A Four Mechanism Model for Threshold Spatial Vision," *Vision Research*, vol. 19, pp. 19–32, 1979.

[74] D. T. Nguyen, "From Human Vision to Image Processing," *The Institution of Professional Engineers New Zealand*, vol. 15, no. 1, pp. 37–47, 1988.

[75] F. W. Mounts, "A video encoding system with conditional picture-element replenishment," *The Bell-System Technical Journal*, vol. 48, pp. 2545–2554, Sept. 1969.

[76] J. C. Candy, "Transmitting television as clusters of frame to frame differences," *The Bell-System Technical Journal*, vol. 50, pp. 1889–1917, Aug. 1971.

[77] B. G. Haskel, F. W. Mounts and J. C. Candy, "Interframe Coding of Videotelephone Pictures," *Proceedings of the IEEE*, vol. 60, pp. 792–800, July 1972.

[78] D. Westerkamp, "Adaptive Intra-/Interframe DPCM-Coding for Transmission Colour TV-Signals with 34 Mbit/s," *IEEE Zurich Seminar on Digital Communication*, pp. 39–45, Mar. 1984.

[79] J. A. Stuller, "Interframe television coding using gain and displacement compensation," *The Bell-System Technical Journal*, vol. 59, pp. 1227–1240, Sept. 1980.

[80] J. O. Limb and J. A. Murphy, "Measuring the speed of moving objects from television signals," *IEEE Transaction on Communication*, vol. 23, pp. 474–478, Apr. 1975.

[81] C. Cafforio and F Rocca, "Methods for Measuring Small Displacements of Television Images," *IEEE Transaction on Information Theory*, vol. 22, pp. 573–579, Sept. 1976.

[82] A. N. Netravali and J. D. Robbins, "Motion Compensated Television Coding : Part 1," *The Bell-System Technical Journal*, vol. 58, pp. 631–670, Mar. 1979.

[83] C. Cafforio and F Rocca, "The Differential Method for Motion Estimation," *in Image Sequence Processing and Dynamic Scene Analysis, T. S. Huang Editor, Springer-Verlag, New York*, pp. 104–124, 1983.

[84] J. D. Robbins and A. N. Netravali, "Recursive Motion Compensation : A Review," *in Image Sequence Processing and Dynamic Scene Analysis, T. S. Huang Editor, Springer-Verlag, New York*, p. 75, 1983.

[85] J. R. Jain and A. K. Jain, "Displacement Measurement and Its Application in Interframe Image Coding," *IEEE Transaction on Communication*, vol. 29, pp. 1799–1808, Dec. 1981.

[86] T. Koga, K. Iinuma, A. Hirano, Y. Iijima and T. Ishiguro, "Motion-compensated interframe coding for video conferencing," *Proc. Nat. Telecommun. Conf., New Orleans, LA.*, pp. G5.3.1–5.3.5, Nov. 1981.

[87] R. Srinivasan and K. R. Rao, "Predictive coding based on efficient motion compensation," *IEEE Int.Conf.Commun., Amsterdam*, pp. 521–526, May 1984.

[88] A. Puri and H. M. Hang and D. L. Schilling, "An efficient block matching algorithm for motion compensated coding," *Proc. IEEE ICASSP*, pp. 25.4.1–25.4.4, 1987.

[89] M. Ghanbari, "The Cross-Search Algorithm for Motion Estimation," *IEEE Transaction on Communication*, vol. 38, pp. 950–953, July 1990.

[90] M. Gotze and G. Ocylock, "An adaptive interframe transform coding system for images," *Proc. IEEE ICASSP 82*, pp. 448–451, 1982.

[91] J. A. Roese, W. K. Pratt and G. S. Robinson, "Interframe Cosine Transform Image Coding," *IEEE Transaction on Communication*, vol. 25, pp. 1329–1338, Nov. 1977.

[92] J. A. Roese, *Hybrid Transform/Predictive Image Coding, in Image Transmission Techniques, W. K. Pratt (Ed.).* Academic Press, New York, 1979.

[93] D. LeGall, "MPEG : A Video Compression standard for multimedia applications," *Communication ACM*, vol. 34, pp. 46–58, Apr. 1991.

[94] A. Haar, "Zur Theorie der Orthogonalen Funktionen Systeme," *Math. Annal.*, vol. 69, pp. 331–337, 1910.

[95] P. Franklin, "A Set of Continuous Orthogonal Functions," *Math. Annal.*, vol. 100, pp. 522–529, 1928.

[96] R. K. Young, *Wavelet theory and its applications, Chapter 2 pp.19-69.* Kluwer publishers, Boston, 1993.

[97] C. K. Chui, "An Introduction to Wavelets," *Academic Press, New York*, 1992.

[98] I. Daubehies, "Orthonormal bases of compactly supported wavelets," *Communication Pure & Applied Mathematics*, vol. 41, pp. 909–996, Nov. 1988.

[99] I. Daubechies, "Ten lectures on wavelets," *Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania*, 1992.

[100] A. Cohen, I. Daubechies and J. C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *AT&T Bell Laboratories, preprint*, 1990.

[101] A. Grossman and Y. Meyer, "Decomposition of Hardy functions in square-integrable wavelets of constant shape," *SIAM Journal on Mathematical Analysis*, vol. 15, pp. 723–736, July 1984.

[102] I. Daubechies, "Time-frequency localization operators : a geometric phase space approach," *IEEE Transaction on Information Theory*, vol. 34, pp. 605–612, 1988.

[103] Y. Meyer, *Orthonormal Wavelets in Wavelets, Time-Frequency Method and Phase Space.* Lecturer Notes on IPTI, J. M. Combes et al., Eds. Springer-Verlag, New York, 1989.

[104] I. Daubechies, "The Wavelet Transform, Time-Frequency Localization and Signal Analysis," *IEEE Transaction on Information Theory*, vol. 36, pp. 961–1005, Sept. 1990.

[105] S. G. Mallat, "Multiresolution Approximation and Wavelet Orthonormal Bases of $L^2(\mathbf{R})$," *Trans. Amer. Math. Soc.*, vol. 315, pp. 69–87, Sept. 1989.

[106] Y. Meyer, "Ondelettes et Operateurs, Tome I.," *Ondelettes, Hermann ed., Paris*, 1990.

[107] G. Beyklin, R. Coifman and V. Rokhlin, "Fast Wavelet Transform and Numerical Algorithms I," *Communication Pure & Applied Mathematics*, vol. 44, pp. 141–183, 1991.

[108] M. Antonini, M. Barlaud, P. Mathew and I. Daubechie, "Image coding using wavelet transforms," *IEEE Transaction on Image Processing*, vol. 1, pp. 205–220, Apr. 1992.

[109] R. R. Coifman and Y. Meyer, "The Discrete Wavelet Transform," *Dept. of Math. Yale Univ., preprint*, 1991.

[110] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser, "Signal Processing and Compression with Wave Packets," *Numerical Algorithms Research Group, New Haven, CT Yale University*, 1990.

[111] M. V. Wickerhauser, "Acoustic Signal Compression with Wavelet Packets," *in Wavelets : A Tutorial in Theory and Applications, C.K. Chui, Ed. New York : Academic, pp. 679-700.*, 1992.

[111] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser, "Signal Processing and Compression with Wave Packets," *Numerical Algorithms Research Group, New Haven, CT Yale University*, 1990.

[112] M. V. Wickerhauser, "Acoustic Signal Compression with Wavelet Packets," *in Wavelets : A Tutorial in Theory and Applications, C.K. Chui, Ed. New York : Academic, pp. 679-700.*, 1992.

[113] CCIR Recomendation 500 - 3, "Methods for the Subjective Assessment of the Quality of Television Pictures," 1986.

[114] D. Gunawan and D.T. Nguyen, "Subjective Coding of Wavelet Coefficients for Multiresolution Analysis," *Proceedings of Image & Vision Computing Conference, Auckland, New Zealand*, pp. 163–168, Aug. 1993.

[115] S. P. Lloyd, "Least Square Quantization in PCM," *IEEE Transaction on Information Theory*, vol. 28, pp. 129–137, Mar. 1982.

[116] J. Max, "Quantizing for Minimum Distortion," *IRE Transaction on Information Theory*, vol. 6, pp. 7–12, 1960.

[117] Jurgen Pandel, "Variable Bit-Rate Image Sequence Coding with Adaptive Quantization," *Signal Processing*, vol. 3, pp. 123–128, June 1991.

[118] D. T. Nguyen, "A Unified Approach to Differential Edge Detectors," *In 23rd New Zealand National Electronic Conference Proceedings, Palmerston North*, pp. 161–167, Aug. 1986.

[119] K. Ramchandran and M. Vetterli, "Best Wavelet Packet Bases in a Rate-Distortion Sense," *IEEE Transaction on Image Processing*, vol. 2, pp. 160–175, Apr. 1993.

[120] R. R. Coifman, Y. Meyer and M. V. Wickerhauser, "Size properties of wavelet packets," *in Wavelets and Their Applications, Edited by Ruskai et.al*, pp. 453–470, 1992.

[121] M. V. Wickerhauser, "Picture compression by best-basis subband coding," *Dept. of Math. Yale Univ., preprint*, 1990.

[122] M. V. Wickerhauser, "Lectures on Wavelet Packet Algorithm," *In Wavelets Down Under Workshop, Adelaide, Australia*, pp. 1–75, Jan. 1993.

[123] A. N. Netravali and J. O. Limb, "Picture coding : A Review," *Proceedings of the IEEE*, vol. 68, pp. 366–406, Mar. 1980.

[124] T. Ishiguro and K. Iinuma, "Television bandwidth compression transmission by motion-compensated interframe coding," *IEEE Communication Magazine*, vol. 10, pp. 24–30, 1982.

[125] H. G. Musmann and P. Pirch and H. J. Grallert, "Advances in Picture Coding," *Proceedings of the IEEE*, vol. 73, pp. 523–548, Apr. 1985.

[126] M. Liou, "Overview of the px64 kbits/s video coding standard," *Communication ACM*, vol. 34, pp. 59–63, Apr. 1991.

[127] Y. Ninomiya and Y. Ohtsuka, "A Motion-Compensated Inter-frame Coding Scheme for Television Pictures," *IEEE Transaction on Communication*, vol. 30, pp. 201–211, Jan. 1982.

[128] S. Kappagantula and K. R. Rao, "Motion Compensated Inter-frame Image Prediction," *IEEE Transaction on Communication*, vol. 33, pp. 1011–1015, Sept. 1985.

[129] M. H. Chan and Y. B. Yu and A. G. Constantinides, "Variable Size Block Matching Motion Compensation with Application to Video Coding," *IEE Proceeding Part I*, vol. 137, pp. 205–212, Aug. 1990.

[130] Q. Wang and R. J. Clarke, "A New Motion Compensated Image Sequence Coding Scheme," *IEE Proceeding Part I*, vol. 39, pp. 219–223, Apr. 1992.

[131] H. Kaneko and T. Ishiguro, "Digital Television Transmission Using Bandwidth Compression Techniques," *IEEE Communication Magazine*, pp. 14–22, July 1980.

[132] T. Koga and M. Ohta, "Entropy Coding for a Hybrid Scheme with Motion Compensation in Subprimary Rate Video Transmission," *IEEE Journal on Selected Areas in Communications*, vol. 5, pp. 1166–1174, Aug. 1987.

[133] S. Zafar and Y. Q. Zhang and B. Jabbari, "Multiscale Video Representation Using Multiresolution Motion Compensation and Wavelet Decomposition," *IEEE Journal on Selected Areas in Communications*, vol. 11, pp. 24–34, Jan. 1993.

[134] K. Uz, M. Vetterli and D. LeGall, "Interpolative multi-resolution coding of advanced television and compatible subchannels," *IEEE Trans. Circ. and Syst. for Video Technol.*, vol. 1, pp. 14–38, Oct. 1991.

[135] Andy Redfern, "A Discrete Transformation," *Australian Personal Computer*, vol. 10, pp. 144–154, Dec. 1989.

[136] D. J. Sakrison, "On the Role of the Observer and a Distortion Measure in Image Transmission," *IEEE Transaction on Communication*, vol. 25, pp. 1251–1267, Nov. 1977.

[137] D. J. Granrath, "The role of human visual models in image processing," *Proceedings of the IEEE*, vol. 69, pp. 552–561, May 1981.

[138] F. X. J. Lukas and Z. L. Budrikis, "Picture quality prediction based on visual model," *IEEE Transaction on Communication*, vol. 30, pp. 1679–1692, July 1982.

[139] M. Miyahara, "Quality assessment for visual service," *IEEE Communication Magazine*, pp. 51–60, Oct. 1988.

[140] J. Katto, K. Onda and Y. Yasuda, "Variable bit-rate coding based on human visual system," *Signal Processing*, vol. 3, pp. 313–320, Sept. 1991.